PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61B

A2

(11) International Publication Number: **WO 99/20168**

(43) International Publication Date: 29 April 1999 (29.04.99)

(21) International Application Number: PCT/NZ98/00160

(22) International Filing Date: 19 October 1998 (19.10.98)

(30) Priority Data:
328994 17 October 1997 (17.10.97) NZ

(71) Applicants *(for all designated States except US)*: UNIVERSITY OF OTAGO [NZ/NZ]; Leith Street, Dunedin (NZ). TE WHETU WHANAU TRUST LIMITED [NZ/NZ]; P.O. Box 4341, Mount Maunganui South (NZ).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: REEVE, Anthony, Edmund [NZ/NZ]; 22 Como Street, Dunedin (NZ). GUILFORD, Parry, John [NZ/NZ]; 140 Queen Street, Dunedin (NZ).

(74) Agents: BENNETT, Michael, Roy et al.; Russell McVeagh West–Walker, The Todd Building, Level 5, 171–177 Lambton Quay, Wellington 6001 (NZ).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: GERMLINE MUTATIONS IN THE E–CADHERIN GENE AND METHOD FOR DETECTING PREDISPOSITION TO CANCER

(57) Abstract

This invention relates to methods by which a predisposition to cancer can be determined. In particular, it relates to methods for detecting whether a patient has a predisposition to cancer, particularly hereditary diffuse gastric cancer with reference to an alteration (mutation) in the gene encoding E–cadherin.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

GERMLINE MUTATIONS IN THE E-CADHERIN GENE AND METHOD FOR DETECTING PREDISPOSITION TO CANCER

This invention relates to methods by which a predisposition to cancer can be determined. In particular, it relates to methods for detecting whether a patient has a predisposition to cancer, particularly hereditary diffuse gastric cancer.

BACKGROUND

The key to cancer treatment is early detection. The ability to predict who is at extreme risk, before the appearance of clinical symptoms, will enable the earliest possible detection of malignancy (watchful waiting). It will also enable prophylactic intervention prior to the onset of clinical signs.

It is therefore the object of this invention to provide a predictive method by which susceptibility to cancer, particularly gastric cancer, can be determined or at least to provide the public with a useful choice.

Gastric cancer remains a major cause of cancer death worldwide, and about 10% of cases show familial clustering. The relative contributions of inherited susceptibility and environmental effects to familial gastric cancer are poorly understood because little is known of the genetic events that predispose to gastric cancer.

The identification of genes predisposing to familial cancer is therefore an essential step towards understanding the molecular events underlying tumourigenesis and is critical for the clinical management of affected families.

The applicants have identified a gene in individuals which, when mutated, predisposes that individual towards developing cancer, particularly hereditary gastric cancer. It is this finding, and the implications it has for cancer screening and management (particularly for families with a history of familial cancer) which underlies the present invention.

SUMMARY OF THE INVENTION

Accordingly, in a first aspect, the invention broadly provides a method of testing to detect whether an individual is predisposed to cancer which comprises the step of detecting the presence or absence of an alteration (mutation) in the gene encoding E-cadherin.

In a further aspect, the invention provides a method of assessing the risk in a human subject for a predisposition for cancer which comprises the step of determining whether there is a germline alteration in the gene encoding E-cadherin, wherein the presence of an alteration is indicative of a risk for a predisposition for cancer.

As used herein gene encoding E-cadherin" means not only the coding sequence for wild-type E-cadherin but also includes non-coding flanking sequences and regulatory elements, mutations in which can cause transcript instability and/or transcriptional repression, and the sites for transcript splicing. These include the two nucleotides immediately upstream (usually "AG") and the two nucleotides immediately downstream (usually "GT") of each exon, and also the splicing branch site located 18-38 bp upstream of each exon.

In one (preferred) embodiment, the presence or absence of the mutation is detected through analysis of the DNA encoding E-cadherin and/or its regulatory elements.

In an alternative embodiment, the presence or absence of the mutation is detected through analysis of mRNA transcribed from the DNA encoding E-cadherin.

In still a further embodiment, the presence or absence of the mutation is detected through analysis of the amino acid sequence of the expressed E-cadherin protein.

As a separate embodiment, the invention provides a method of prophylaxis and/or therapeutic treatment against cancer of an individual identified as having a risk of predisposition to cancer by a method defined above which comprises the step of

increasing, maintaining and/or restoring the active concentration of wild-type E-cadherin protein within said individual.

Conveniently, the method will be a gene therapy method and will involve supplying the individual with wild-type E-cadherin gene function.

DESCRIPTION OF THE DRAWINGS

While the invention is broadly as defined above, it will be appreciated that it is not limited thereto but that it also includes embodiments of which the following description provides examples. In addition, the invention will be better understood through reference to the accompanying drawings in which:

Figure 1 shows the nucleotide and amino acid sequences for wild-type E-cadherin cDNA;

Figure 2 is a kindred map for one family (family A) having a predisposition to gastric cancer. Numbers to the right of the symbols indicate age at death. The age is underlined if a blood or biopsy sample was available. General symbols: squares, males; circles, females; all symbols with a diagonal, deceased. Solid symbols: gastric carcinoma, pathology available; dotted symbols: gastric carcinoma, pathology unavailable; vertical stripes: colorectal cancer;

Figure 3 is a graph showing the age of death from gastric cancer in the studied kindred of family A;

Figure 4 shows the results of a mutation analysis of exon 7 of the E-cadherin gene as follows:

(a) SSCP pattern of exon 7 in E-cadherin gene. The SSCP band pattern of two affected people, two obligate carriers and two unaffected spouses (wild type) are shown. The additional band in the affected and obligate carrier samples is indicated by the arrow;

(b) Direct sequence analysis of the exon-intron boundary of exon 7 showing the wild type sequence and the sequence from an affected person heterozygous for the G to T transversion. The position of the exon/intron boundary is marked;

Figure 5 is an abbreviated kindred map for a second family (family B). General symbols: squares, males; circles, females; all symbols with a diagonal, deceased. Solid symbols: gastric carcinoma, pathology available; dotted symbols: gastric carcinoma, pathology unavailable; vertical stripes: colorectal cancer. Diagonal hatching: unconfirmed gastric carcinoma;

Figure 6 shows sequence analysis results for DNA from family B (Figure 6A) and family C (Figure 6B), exons 15 and 13 respectively;

Figure 7 shows pedigrees of non-Maori gastric cancer families. General symbols: squares, males; circles, females; all symbols with a diagonal, deceased. Patient numbers are included; and

Figure 8 shows mutations in gastric cancer families. **(a).** Exon 11 DNA sequence from family 1000 showing the insertion of an additional C nucleotide between the G at position 1588 and the A at position 1591. **(b).** Exon 2 sequence from family 4201 showing the heterozygous (G/T) mutation at position 70. **(c).** Exon 8 / intron 8 sequence of family CHG 72 showing the heterozygous (G/A) mutation at the first nucleotide of the intron. Nucleotide positions are as described in Berx *et al.* (1995). Sequencing products were analysed on a LiCor 4000L DNA sequencer.

DESCRIPTION OF THE INVENTION

As defined above, the method of the invention detects a predisposition to cancer. The critical finding made by the applicants is that this predisposition is due to an

alteration (mutation) in the gene encoding E-cadherin. This finding forms the basis of the present invention.

E-cadherin is a transmembrane protein with five tandemly repeated extracellular domains and a cytoplasmic domain which connects to the actin cytoskeleton via a complex with α, β and γ catenins (Grunwald (1993)). It plays an important role in establishing cell polarity and maintaining normal tissue morphology and cellular differentiation. Diminished E-cadherin expression is associated with poorly differentiated carcinomas which display aggressive histopathologic characteristics such as infiltrative growth and lymph node involvement (Shiozaki *et al.* (1995)). Under-expression has been proposed as a prognostic marker of poor clinical outcome in many tumour types (Bracke *et al.* (1996)). In experimental tumour models, restored expression of E-cadherin can suppress the invasiveness of epithelial tumour cells (Frixen (1991), Vlemincke (1991)).

However, to date, there has been no suggestion that an alteration/mutation in the gene encoding E-cadherin is in any way predictive of susceptibility to cancer prior to tumourigenesis.

The amino acid and cDNA nucleotide sequences encoding wild-type E-cadherin are shown in Figure 1. Any change in either sequence is included in the scope of the term "mutation" as used herein.

The gene encoding E-cadherin was identified as a susceptibility gene through genetic linkage analysis. This analysis was performed in relation to samples obtained from a large (Maori) kindred from New Zealand, the pedigree pattern of which is shown in Figure 2 (family A). This pedigree pattern is consistent with the dominant inheritance of a susceptibility gene with incomplete penetrance.

The linkage analysis determined that the susceptibility to cancer was associated with the gene encoding E-cadherin. This was confirmed with reference firstly to two further Maori kindreds (families B and C) and then to non-Maori kindreds.

In one approach, according to the present invention, alteration of the wild-type E-cadherin gene is detected. In addition, the method can be performed by detecting the wild-type E-cadherin gene and confirming the lack of a predisposition or neoplasia.

"Alteration of a wild-type E-cadherin gene" encompasses all forms of mutations including deletions, insertions and point mutations in the coding and noncoding regions. Deletions may be of the entire gene or only a portion of the gene. Point mutations may result in stop codons, frameshift mutations or amino acid substitutions.

The alterations or mutations which are focus of the predictive method of the invention are germline mutations. Germline mutations can be found in any of a body's tissues and are inherited.

Mutations leading to non-functional gene products primarily lead to a cancerous state. However, mutations which lead to decreased expression of the E-cadherin gene product will also lead to a cancerous condition. Point mutation events may occur in regulatory regions, such as in the promoter of the gene, leading to loss or diminution of expression of the mRNA. Point mutations may also abolish proper RNA processing, leading to loss of expression of the E-cadherin gene product, or a decrease in mRNA stability or translation efficiency.

Predisposition to cancers, such as diffuse gastric cancer and the other cancers identified herein, can be ascertained by testing any tissue of a human for mutations of the E-cadherin gene. For example, a person who has inherited a germline E-cadherin mutation would be prone to develop cancers. This can be determined by testing DNA from any sample from the person's body such as serum, sputum and urine. Most simply, blood can be drawn and DNA extracted from the cells of the blood. In addition, prenatal diagnosis can be accomplished by testing fetal cells, placental cells or amniotic fluid for mutations of the E-cadherin gene.

A preliminary analysis to detect deletions in DNA sequences can be performed by looking at a series of Southern blots of DNA cut with one or more restriction

enzymes, preferably a large number of restriction enzymes. Each blot contains DNA from a series of normal individuals and from a series of test cases. Southern blots displaying hybridizing fragments (differing in length from control DNA when probed with sequences near or including the E-cadherin locus) indicate a possible mutation. If restriction enzymes which produce very large restriction fragments are used, then pulsed field gel electrophoresis ("PFGE") can be employed.

Detection of point mutations may be accomplished by molecular cloning of the E-cadherin allele(s) and sequencing that allele(s) using techniques well known in the art. Alternatively, the gene sequences can be amplified, using known polynucleotide amplification techniques, directly from a genomic DNA preparation from the sample tissue. The amplification techniques which can be used include methods such as the polymerase chain reaction (PCR), ligation amplification (or ligase chain reaction, LCR) and amplification methods based on the use of Q-beta replicase. These methods are well known and widely practised in the art. See, eg., US Patents 4,683,195 and 4,683,202 and Innis *et al.*, 1990 (for PCR); and Wu *et al.*, 1989a (for LCR). Reagents and hardware for conducting amplification are commercially available. Primers useful to amplify sequences from the E-cadherin region are preferably complementary to, and hybridize specifically to sequences in the E-cadherin region or in regions that flank a target region therein.

E-cadherin sequences generated by amplification may be sequenced directly. Alternatively, but less desirably, the amplified sequence(s) may be cloned prior to sequence analysis. A method for the direct cloning and sequence analysis of enzymatically amplified genomic segments has been described by Scharf, 1986.

There are numerous well known methods for confirming the presence of a susceptibility allele. These include: 1) single stranded confirmation analysis ("SSCA") (Orita *et al.*, 1989); 2) denaturing gradient gel electrophoresis ("DGGE") (Wartell *et al.*, 1990; Sheffield *et al.*, 1989); 3) RNase protection assays (Finkelstein *et al.*, 1990; Kinsler *et al.*, 1991); 4) allele-specific oligonucleotides (ASO's) (Conner *et al.*, 1983); 5) the use of proteins which recognize nucleotide mismatches, such as the *E. coli* mutS protein (Modrich, 1991); and 6) allele-specific PCR (Rano & Kidd, 1989). For allele-specific PCR, primers are used which hybridize at their 3' ends to a

particular E-cadherin mutation. If the particular E-cadherin mutation is not present, an amplification product is not observed.

Other approaches which can also be used include the Amplification Refractory Mutation System (ARMS), as disclosed in European Patent Application Publication No. 0332435 and in Newton *et al.*, 1989. Insertions and deletions of genes can also be detected by cloning, sequencing and amplification. In addition, restriction fragment length polymorphism (RFLP) probes for the gene or surrounding marker genes can be used to detect alteration of an allele or an insertion in a polymorphic fragment. Such a method is particularly useful for screening relatives of an affected individual for the presence of the E-cadherin mutation found in that individual.

In the first three methods (ie., SSCA, DGGE and RNase protection assay), a new electrophoretic band appears. SSCA detects a band which migrates differentially because the sequence change causes a difference in single-strand, intramolecular base pairing. RNase protection involves cleavage of the mutant polynucleotide into two or more smaller fragments. DGGE detects differences in migration rates of mutant sequences compared to wild-type sequences, using a denaturing gradient gel. In an allele-specific oligonucleotide assay, an oligonucleotide is designed which detects a specific sequence, and the assay is performed by detecting the presence or absence of a hybridization signal. In the mutS assay, the protein beings only to sequences that contain a nucleotide mismatch in a heteroduplex between mutant and wild-type sequences.

Mismatches are hybridized nucleic acid duplexes in which the two strands are not 100% complementary. Lack of total homology may be due to deletions, insertions, inversions or substitutions. Mismatch detection can be used to detect point mutations in the gene or its mRNA product. While these techniques are less sensitive than sequencing, they are simpler to perform on a large number of samples.

An example of a mismatch cleavage technique is the RNase protection method. This method involves the use of a labeled riboprobe which is complementary to the human wild-type E-cadherin gene coding sequence. The riboprobe and either

mRNA or DNA isolated from the test tissue are annealed (hybridized) together and subsequently digested with the enzyme RNase A which is able to detect some mismatches in a duplex RNA structure. If a mismatch is detected by RNase A, it cleaves at the site of the mismatch. Thus, when the annealed RNA preparation is separated on an electrophoretic gel matrix, if a mismatch has been detected and cleaved by RNase A, an RNA product will be seen which is smaller than the full length duplex RNA for the riboprobe and the mRNA or DNA.

The riboprobe need not be the full length of the E-cadherin mRNA or gene but can be a segment of either. If the riboprobe comprises only a segment of the E-cadherin mRNA or gene, it will be desirable to use a number of these probes to screen the whole mRNA sequence for mismatches.

In similar fashion, DNA probes can be used to detect mismatches, through enzymatic or chemical cleavage. See, eg., Cotton *et al.,* 1989; Shenk *et al.,* 1975; Novack *et al.,,* 1986. Alternatively, mismatches can be detected by shifts in the electrophoretic mobility of mismatched duplexes relative to matched duplexes. See eg. Cariello, 1988. With either riboprobes or DNA probes, the cellular mRNA or DNA which might contain a mutation can be amplified using PCR before hybridization. Changes in DNA of the E-cadherin gene can also be detected using Southern hybridization, especially if the changes are gross rearrangements, such as deletions and insertions.

DNA sequences of the E-cadherin gene which have been amplified by use of PCR may also be screened using allele-specific probes. These probes are nucleic acid oligomers, each of which contains a region of the E-cadherin gene sequence harboring a known mutation. For example, one oligomer may be about 30 nucleotides in length, corresponding to a portion of the E-cadherin gene sequence. By use of a battery of such allele-specific probes, PCR amplification products can be screened to identify the presence of a previously identified mutation in the E-cadherin gene.

Hybridization of allele-specific probes with amplified E-cadherin sequences can be performed, for example, on a nylon filter such as Hybond. Hybridization to a

particular probe under stringent hybridization conditions indicates the presence of the same mutation in the tumour tissue as in the allele-specific probe.

Mutations from potentially susceptible patients falling outside the coding region of E-cadherin can be detected by examining the non-coding regions, such as introns and regulatory sequences near or within the E-cadherin gene. An early indication that mutations in noncoding regions are important may come from Northern blot experiments that reveal messenger RNA molecules of abnormal size or abundance in cancer patients as compared to control individuals.

Alteration of E-cadherin mRNA expression can be detected by any techniques known in the art. These include Northern blot analysis, PCR amplification and RNase protection. Diminished mRNA expression indicates an alteration of the wild-type E-cadherin gene. Alteration of wild-type E-cadherin genes can also be detected by screening for alteration of wild-type E-cadherin protein. For example, monoclonal antibodies immunoreactive with wild-type E-cadherin can be used to screen a tissue with lack of bound antigen indicating an E-cadherin mutation.

Monoclonal antibodies with affinities of 10^{-8} M^{-1} or preferably 10^{-9} to 10^{-10} M^{-1} or stronger will typically be made by standard procedures as described, eg. in Harlow & Lane, 1988 or Goding, 1986. Briefly, appropriate animals will be selected and the desired immunization protocol followed. After the appropriate period of time, the spleens of such animals are excised and individual spleen cells fused, typically, to immortalised myeloma cells under appropriate selection conditions. Thereafter, the cells are clonally separated and the supernatants of each clone tested for their production of an appropriate antibody specific for the desired region of the antigen.

Other suitable techniques for preparing antibodies involve *in vitro* exposure of lymphocytes to the antigenic polypeptides, or alternatively, to selection of libraries of antibodies in phage or similar vectors. See Huse *et al.*, 1989.

Also, recombinant immunoglobulins may be produced using procedures known in the art (see, for example, US Patent 4,816,567).

The antibodies may be used with or without modification. Frequently, antibodies will be labeled by joining, either covalently or non-covalently, a substance which provides for a detectable signal. A wide variety of labels and conjugation techniques are known and are reported extensively in the literature. Suitable labels include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent agents, chemiluminescent agents, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241.

Antibodies specific for products of mutant alleles could also be used to detect mutant E-cadherin gene product. Such antibodies can be produced in equivalent fashion to the antibodies for wild-type E-cadherin as described above.

The immunological assay in which the antibodies are employed can involve any convenient format known in the art. Such formats include Western blots, immunohistochemical assays and ELISA assays. In addition, functional assays such as protein binding determinations, can also be used.

In summary, any approach to detecting a germline alteration in the underlying DNA coding for wild-type E-cadherin expression can be employed, whether the analysis be of the DNA itself, mRNA transcribed from the DNA or the protein which is the ultimate expression product of the DNA.

The following experimental sections outline the various analyses undertaken in detail. These identify a number of different mutations and are included for reasons of exemplification only.

EXPERIMENTAL

SECTION 1 - Familial gastric cancer in Maori kindreds (families A, B and C)

Methods

Genotyping: DNA extracted from blood and biopsy samples (Banerjee *et al.*, (1995)) was genotyped using standard conditions (Dib, (1996)) in reactions containing 0.2U

AmpliTaq Gold (Perkin Elmer) and 25 pmole of infrared labelled (IR41) forward primer (MWG Biotech). Products were analysed on a LiCor 4000L DNA sequencer.

SSCP analysis: SSCP mutation analysis was carried out as described by Berx *et al.*, (1995). The PCR products were electrophoresed at room temperature through a 6% non denaturing polyacrylamide gel without added glycerol. Products were detected by autoradiography.

RT-PCR: Total RNA was extracted (Chomczynski *et al.*, (1987)) from frozen biopsy material and reverse transcribed using SuperScript II (Gibco BRL) according to the manufacturers instructions. Nucleotide position 1008 was PCR-amplified from the cDNA using a forward primer within exon 7 (5'-TAA CAG GAA CAC AGG AGT CAT CA-3') and a reverse primer from exon 8 (5'-GTG GTG GGA TTG AAG ATC GG-3'). Reactions contained 4mM $MgCl_2$ and 0.2U AmpliTaq Gold and were cycled as follows: (95°C 10 min) 1 cycle and (95°C 15 sec, 57°C 45 sec, 72°C 10 sec) for 35 cycles.

Plasmid and direct sequencing: RT-PCR products were eluted from a 6% polyacrylamide denaturing gel, re-amplified with the original primers using *Pwo* polymerase (Boehringer Mannheim) and ligated into the *Eco*RV site of Bluescript. Template for direct sequencing of mutations was produced from genomic DNA by PCR using the SSCP antisense primers and the sense primers[8] (Berx *et al.*, 1995) with an added 5' leader corresponding to the T3 sequencing primer. Plasmid and direct sequencing were carried out using Thermosequenase (Amersham) and an IR41 labelled (MWG Biotech) T3 primer (3 pmoles/reaction). The products were analysed on a LiCor 4000L DNA sequencer.

Linkage analysis: Two point lod scores were calculated using MLINK of the LINKAGE 5.1 package (Lathrop *et al.*, (1985)). A gene frequency of 10^{-4} was assumed for the disease gene. Age dependent penetrance was taken into account; seven liability classes were obtained from the cumulative age of onset curve: 0.18 for individuals from 0-20yrs, 0.24 (21-25 years), 0.34 (26-30 years), 0.48 (31-35 years), 0.56 (36-40 years), 0.64 (41-45 years) and 0.70 (>46 years). Variation of the maximum penetrance from 60-80% did not change the significance of the results.

Results

Linkage analysis

Reference should be made to Table 1 below which relates to family A.

Table 1. Two point lod scores for linkage of the gastric cancer susceptibility gene to markers mapping to the genetic interval containing E-cadherin

Marker	Lod scores		Recombination fraction (q)
	Equal allele frequencies	Kindred allele frequencies	
D16S752*	5.04	4.04	0
D16S3043**	2.01	2.34	0.05
D16S3019**	2.28	1.57	0
D16S3095**	4.90	4.07	0
D16S3083**	2.79	2.16	0
D16S3138**	3.32	2.68	0

Lod scores were calculated assuming either equal allele frequencies or, in a conservative approach (in the absence of allele frequencies for the study population), using the actual frequencies observed in the study kindred.

* GDB (TM) Human Genome Database, Baltimore (Maryland, USA): John Hopkins University.

** Dib, 1996.

The linkage analysis found a maximum two-point lod score (Zmax=5.04, θ=0) with marker D16S752, which maps within the genetic interval on chromosome 16q22.1 containing the E-cadherin gene (GDB, Human Genome Database, Baltimore, Maryland, USA; John Hopkins University. Genotyping of five other markers (Dib, (1996)) in the vicinity of E-cadherin identified additional significantly linked markers. A conserved haplotype spanning 9 centimorgans from D16S3019 to D16S3138 was consistently inherited with the disease. This haplotype was also present in all obligate carriers of the susceptibility gene and a proportion of the unaffected individuals. The proportion of individuals with this haplotype who were affected by the age of 60 provided an approximation of 70% for the penetrance of the susceptibility gene in this kindred.

Mutation analysis - family A

Mutation analysis of samples from the kindred of Figure 2 (family A) using the single-stranded conformational polymorphism (SSCP) technique (Berx *et al.*, 1995) revealed a band-shift in exon 7 (Fig. 4a) in DNA extracted from lymphocytes of two affected people and four obligate carriers of the susceptibility gene. Direct sequencing of exon 7 identified a G-T transversion at the last nucleotide (position 1008) of this exon (Fig. 4b). The SSCP band-shift was not observed in 150 unrelated chromosomes (data not shown).

Discussion - exon 7 mutation

The consequences of the mutation G-T transversion in exon 7 are two-fold. Firstly, the mutated nucleotide forms part of the splice donor site consensus sequence (Padgett *et al.*, (1986)). Mutation of this splice site position results in exon skipping and the activation of cryptic splice sites (Andrews *et al.*, (1982); KuiVaniem *et al.*, (1995)). Mutation of E-cadherin nucleotide 1008 (a G to A transition) has been observed previously in a cell line derived from a histologically diffuse gastric carcinoma (Kato III) (Oda *et al.*, (1994)). This mutation resulted in the activation of cryptic splice sites which led to premature chain termination. To determine the extent to which transcript carrying the G to T transversion was incorrectly spliced, exon-linking RT-PCR (exons 7-8) was performed on stomach biopsy material taken from an affected family member. In addition to the expected product of 180 bp, a minor 187 bp band was also observed. Both products were cloned and resulting clones sequenced. 10/10 clones derived from the larger band contained the mutation and a 7 bp insertion of intronic DNA. The insertion is a consequence of splicing at a cryptic splice site (Oda *et al.*, (1994)). Since transcript which is incorrectly spliced at exon 7 is unstable *in vivo*, the extent of aberrant splicing was estimated from the proportion of correctly spliced transcript which contained the G to T mutation. 1/14 clones derived from the 180 bp product contained the mutation. This result demonstrates that, relative to the wild-type transcript, only about 15% of the mutant transcript accumulates in stomach tissue.

The second consequence of the G-T transversion is the substitution of Glu 336 with Asp (Berx *et al.*, (1995)). Glu 336 is located in one of the LDRE motifs which form

part of E-cadherin's four calcium binding pockets. Calcium binding is required for dimerisation and rigidification of E-cadherin and provides protection from proteolytic degradation (Nagar *et al.*, (1996)). Molecular modelling indicates that an Asp at position 336 would cause a significant deformation in the calcium binding pocket with a probable negative effect on its ability to bind calcium (data not shown). The fact that the LDRE motif is conserved, not only amongst vertebrates but also in Drosophila (Mahoney *et al.*, (1991)), suggests that a Glu to Asp mutation at this position is not tolerated.

Mutation analysis - confirmatory (families B and C)

To confirm the role of E-cadherin in inherited gastric cancer susceptibility, germline mutations in this gene were searched for in two other Maori families (families B and C) with early-onset, histologically diffuse gastric cancer. SSCP analysis of exons 2-16 amplified from lymphocyte DNA was carried out on two affected individuals and one obligate carrier from family B (Fig. 5) and the proband of family C. A band shift was observed in exon 15 in the three members of family B who were tested. Direct sequencing of exon 15 showed that all three individuals were heterozygous for the insertion of an additional C residue in a run of five cytosines at positions 2,382-2,386 (Fig. 6A). The resulting frameshift leads to an E-cadherin molecule lacking about half of its cytoplasmic domain.

The proband of family C (aged 30 years) showed an SSCP band in exon 13. Direct sequencing identified a heterozygous C → T transition at nucleotide 2,095 which converted Gln 699 to a TAG stop codon (Fig. 6B). This inactivating mutation would result in an expressed E-cadherin peptide lacking both the transmembrane and cytoplasmic domain.

Mutation Summary - families A, B and C

The exemplary mutations identified to date in the three Maori kindreds are summarised in Table 2. In addition to the inactivating mutations in families A, B and C, two silent mutations and one missense mutation which did not segregate with the phenotype were found (Table 2).

Table 2. E-cadherin germline mutations and polymorphisms in Maori gastric cancer families

Family	Nucleotide position (exon)	Mutation	Type
A	1,008 (7)	G → T	Splice site
B	2,382-2,386 (15)	C insertion	Frameshift
C	2,095 (13)	C → T	Premature Termination (TAG)
B	1,409 (10)*	C → T	Codon 470: Thr → Ile
A, C	intron 12†	C → T	Silent
A, B, C	2,076 (13)	C → T	Silent

* This mutation did not segregate with the disease in family B.
† Located 13 nucleotides upstream of the exon.

SECTION 2 - Familial gastric cancer in non-Maori kindreds

Material and Methods

Description of families

Family 1000 is of mixed Northern European ancestry (Fig. 7a). The proband and her mother were both diagnosed with high grade adenocarcinoma with signet ring histology and linitis plastica at ages 40 and 48, respectively. The proband's maternal grandfather had died of cancer of unknown type at age 45. A maternal aunt was diagnosed at age 59 with a scirrhous adenocarcinoma of the left breast. At age 63 she also had resection of an adenocarcinoma of the cardia of the stomach. Microscopic examination of the gastric tumour showed a diffuse, poorly differentiated mucous producing adenocarcinoma with numerous signet ring cells.

Family 4201 (Fig. 7b) is of European origin. The family has a strong history of gastric and breast cancer and leukemia. Pathology specimens were available from three of four individuals affected by gastric cancer (III-1, III-2, III-5). These three cancers were all diffusely infiltrative signet ring adenocarcinomas (Watanabe *et al.,* (1990)). Extensive thickening of the stomach wall, consistent with linitis plastica, was

described in one case (III-1). The age at diagnosis of gastric cancer in this family ranged from 37 to 46 years and the age at death ranged from 39 to 55 years. One obligate carrier is unaffected by cancer at age 71 years. However, her sister (II-2) was diagnosed with gastric cancer at age 37 and breast cancer two years later. Two cases of breast cancer alone, and one of Kaposi's sarcoma in the brain (associated with HIV infection) have occurred in this family, with ages at diagnosis of 39, 46 and 40 years, respectively. The histology of these tumours was unavailable. In addition, three family members had unspecified leukemia diagnosed at ages 66, 45, and 45 years. A fourth case of leukemia occurred in a spouse at age 83 years.

Family CHG 72 is of African American origin and has had four family members affected by gastric cancer. The age of diagnosis of the cancers was 25 to 58 with the patients dying between ages 29 and 58. The tumours were all diffuse, poorly differentiated infiltrative adenocarcinomas with signet ring histology. In addition to these four cases, a half sister (II-1) to the proband died of an unconfirmed cancer in her thirties and a child (IV-1) currently aged 10 years suffers from aplastic anemia. The father of the four affected siblings (I-1) died at age 74 of an unknown illness.

DNA manipulation

DNA was extracted from blood using either standard techniques or the Puregene kit (Gentra Systems, Minneapolis, Minnesota) following the manufacturer's protocol. DNA extractions from paraffin-embedded, formalin-fixed tissue were carried out using previously reported techniques (Greer *et al.*, (1995); Grady *et al.*, (1998)). All tumours from family 4201 and family CHG 72 were microdissected prior to DNA extraction. PCR products for the 16 E-cadherin exons were amplified using 1U AmpliTaq Gold (Perkin Elmer) and the primers and conditions described by Berx *et al.* (1996). A 5' leader corresponding to the T3 sequencing primer was added to the sense primer. Direct sequencing of PCR products was carried out using Thermosequenase (Amersham) and an IR800 labelled (MWG Biotech) T3 primer (3 pmoles/reaction). The products were analysed on a LiCor 4000L DNA sequencer.

Confirmation of the E-cadherin mutations in family 4201 and family CHG 72 was performed on PCR products from the genomic DNA extracted from lymphocytes or microdissected, paraffin embedded tumour tissue using the Amplicycle kit (Perkin

Elmer) with aP^{33}-dCTP random priming. To improve the efficiency of PCR amplification of exon 2 when using microdissected tumour DNA, the exon 2 primers were redesigned to amplify a shorter PCR product which contained the region of interest. These primers, 5'-TTC CCC CAC CCC AGG TCT C-3' (EX2F) and 5'-CCC TCA CCT CTG CCC AGG AC-3' (EX2R), correspond to nucleotides 1-19 and 136-117 of the exon 2 genomic sequence (accession # L34937), respectively. Sequencing was performed using either EX2F or the primer 5'-TGT AGC TCT CGG CGT CAA AG-3' (complementary to nucleotides 93-112 of the E-cadherin cDNA sequence (Berx *et al*, 1995)). The sequencing products were electrophoresed on a 6% polyacrylamide 7M urea gel at 70W (50°C) for 50-90 minutes and visualized using either autoradiography or a Storm 820 Phosphorimager (Molecular Dynamics).

Results

Mutation Searching

All 16 E-cadherin exons were PCR amplified and sequenced. Sequencing of peripheral white blood cell DNA from the proband of family 1000 identified the heterozygous insertion of an additional cytosine after position 1588 (1588insC) in exon 11 (Fig. 8a). This frameshift mutation is predicted to lead to premature translation termination in exon 11. The truncated peptide would lack both one third of the extracellular domain and the entire intracellular domain of the wild-type E-cadherin protein. The heterozygous 1588insC mutation was also identified in DNA from the proband's mother who had gastric cancer. This DNA had been extracted from a biopsy of a metastasis to the diaphragm. The biopsy consisted of a mixture of diffusely infiltrating tumour cells and normal stroma.

Sequencing genomic DNA from peripheral white blood cells of the proband of family 4201 (II-1) identified a heterozygous G->T transversion at nucleotide 70 (70G->T) in exon 2 (Fig. 8b). The proband is unaffected but is an obligate carrier of the predisposing mutation. The mutation would convert a glutamic acid (Glu24) to a TAG stop codon in the signal peptide of the E-cadherin precursor protein. This mutation was also identified in microdissected normal tissue from gastric biopsies of three siblings (III-1, III-2, III-5) with gastric cancer, and peripheral white blood cell DNA from an unaffected sibling (III-4) and a first cousin (III-7) affected by breast

cancer. DNA from blood of one unaffected family sibling (III-6) showed no mutation. No other biological samples were available from any of the other family members.

The E-cadherin gene was PCR amplified using peripheral white blood cell DNA from the proband of family CHG 72 (II-4). Sequencing identified a heterozygous G->A transition in the splice donor site of intron 8 (1137+1G->A). Guanine at the +1 position of the splice consensus sequence is 100% conserved in eukaryotic splice sites (Padgett *et al.,* (1986)). The G->A change would be predicted to result in either skipping of exon 8 or the activation of cryptic splice sites. This mutation was identified in DNA from normal and microdissected tumour tissue from paraffin blocks in three additional affected family members (II-2, II-3, III-1). Loss of heterozygosity (LOH) analysis using the microsatellite repeat markers D16S3138, D16S3019, and D16S3043 was performed on the microdissected tumours from family CHG 72 and family 4201 but failed to show LOH.

In addition to the three mutations identified in these families, two silent polymorphisms were also identified in family 1000. The mutations and polymorphisms are summarized in Table 3.

TABLE 3. Summary of mutations and polymorphisms identified in diffuse gastric cancer families.

Family	Mutation/ Polymorphism	Type	Location
1000	1588insC	Frameshift	exon 11
4201	70G->T (E24X)	Nonsense	exon 2
CHG 72	1137+1G->A	Donor splice site	intron 8
1000	2076C->T	Silent	exon 13
1000	1937-27T->G	Silent	intron 12

Table 3. Mutations and polymorphisms were identified by direct sequencing of the 16 E-cadherin exons. Nucleotide positions are as described in Berx *et al,*. (1995).

Discussion

Inactivating germline mutations in the E-cadherin gene have been identified in three of three US families of Caucasian and African American descent with histories consistent with an autosomal dominant susceptibility to diffuse gastric cancer (Lauren (1965)). Tumours in each of the three families were histologically defined as signet ring adenocarcinomas (Watanabe *et al.*, (1990)). These results, taken with the earlier identification of E-cadherin germline mutations in three of three New Zealand Maori families as reported in Section 1 (who are Polynesian in origin) demonstrate that mutation of E-cadherin is a widespread determinant of inherited susceptibility to diffuse gastric cancer, and its occurrence is independent of ethnic origin. Germline E-cadherin mutation therefore genotypically defines an inherited cancer syndrome. This syndrome is designated herein as hereditary diffuse gastric cancer (HDGC).

The high incidence of early-onset breast cancer and unspecified leukemia in family 4201 suggests that non-gastric malignancies may also be associated with HDGC. In addition, one presumed gene carrier in family 1000 had breast cancer prior to developing gastric cancer. E-cadherin mutations have been described in over 50% of sporadic lobular breast cancers (but not in other histopathological subtypes) (Berx *et al.*, (1996)), suggesting that mutation of the E-cadherin gene is required for the onset or progression of this type of cancer. It is notable that of the six families with E-cadherin mutations described above only one (#4201) has an extensive history of cancer at sites other than the stomach. Members of that family carry truncating mutations in the sequence encoding either the E-cadherin signal peptide or the precursor sequence. The remaining mutations would be predicted to result in truncated proteins containing at least part of the extracellular domain including the HAV motif required for E-cadherin homophilic adhesion. Decapeptides containing this motif are capable of inhibiting E-cadherin-mediated cell adhesion (Blaschuk *et al.*, (1990)).

INDUSTRIAL APPLICATION

The above results demonstrate the role that germline mutations in the gene encoding E-cadherin play in susceptibility to cancer, particularly HDGC. Further,

the high frequency of inactivation of the E-cadherin gene in many types of sporadic tumours (Mareel *et al.*, (1995)) suggests that mutations in this gene may also confer inherited susceptibility to other cancers. These include cancers of the breast, prostate, thyroid, liver, kidney, bladder and colon.

The demonstration that mutations in the gene encoding E-cadherin are predictive of cancer susceptibility has a number of implications. As indicated above, the primary implication is in a method of detection of a risk of a predisposition to cancer.

Early at-risk determination provides the opportunity for early intervention. Carriers of the mutation could choose to have prophylactic surgery or chemopreventative treatment prior to the appearance of any malignancy. Testing also enables carriers to make important life decisions (eg. child bearing) and will provide the opportunity for pre-natal diagnosis. For non-carriers, testing will bring peace of mind and will remove the need for surveillance.

The present invention will therefore mean that people from families with histories of familial cancer (such as HDGC) will be able to undergo tests which will search for the presence of E-cadherin gene mutations.

The identification of E-cadherin as a cancer susceptibility gene has implications beyond early detection. The possibility of chemopreventative approaches to delay the onset of cancer is also raised. These approaches, which are based on the activity of the second copy of the E-cadherin gene, fall into two categories: (a) procedures to maintain the expression of the remaining normal E-cadherin gene and (b) procedures to minimise the risk of mutation or loss of the normal allele.

(a). Other than mutation, a number of mechanisms for down-regulating E-cadherin expression are known. These mechanisms are either normal physiological responses such as occur during wound repair or may be consequences of a disease process, as is suggested by the hypermethylation of the E-cadherin gene in a proportion of sporadic tumours. There is also evidence suggesting that E-cadherin can be stored in the cell, possibly in an inactive form. Activation of one or more of these pathways in a person already carrying a mutation in the gene may diminish the concentration

of E-cadherin below the minimum threshold to maintain normal cell adhesion. Since tumourigenesis is a multi-step pathway, under-expression of E-cadherin in a cell which has already acquired mutations in other tumour suppressor genes or oncogenes will accelerate the onset of disease.

Compounds which can increase the expression, or prevent the decrease, of E-cadherin would be potential cancer chemopreventative agents for carriers of mutations in this gene. A number of chemicals are already known to up-regulate E-cadherin:

Insulin-like growth factor-1
9-*cis*-retinoic acid and all-*trans*-retinoic acid
tangeretin
tamoxifen
γ-linolenic acid
calcium
relaxin
17-β estradiol

Alternatively, compounds which prevent wounding in the stomach, such as anti-ulcer treatments, would be predicted to have a protective effect.

(b). Preventing loss of the second E-cadherin allele or other genes involved in the pathway to tumourigenesis will delay the onset of cancer in carriers. Tissue which is inflamed, or undergoing rapid regeneration is more likely to acquire a mutation. Treatments which prevent inflammation or the need for tissue repair should have a protective effect. Therefore compounds which prevent gastritis, antibiotics which eradicate the bacteria *Helicobacter pylori* (which causes inflammation and tissue damage), and anti-ulcer treatments would all offer protection from additional mutations.

There is also the possibility of a curative or corrective approach using gene therapy. This will involve supplying wild-type E-cadherin function to an individual who carries mutant E-cadherin alleles. Supplying such a function should suppress

neoplastic growth of the recipient cells. The wild-type E-cadherin gene or a part of the gene may be introduced into cells within such an individual in a vector such that the gene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. If a gene portion is introduced and expressed in a cell carrying a mutant E-cadherin allele, the gene portion should encode a part of the E-cadherin protein which is required for non-neoplastic growth of the cell. More usual is the situation where the wild-type E-cadherin gene or a part thereof is introduced into the mutant cell in such a way that it recombines with the endogenous mutant E-cadherin gene present in the cell. Such recombination requires a double recombination event which results in the correction of the E-cadherin gene mutation. Vectors for introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector may be used. Methods for introducing DNA into cells such as electroporation, calcium phosphate co-precipitation and viral transduction are known in the art. Cells transformed with the wild-type E-cadherin gene can be used as model systems to study cancer remission and drug treatments which promote such remission.

As generally discussed above, the wild-type E-cadherin gene or fragment, where applicable, may be employed in gene therapy methods in order to increase the amount of the expression products of such genes in cancer cells. Such gene therapy is particularly appropriate for use in pre-cancerous cells, in which the level of E-cadherin polypeptide is absent or diminished compared to normal cells. It may also be useful to increase the level of expression of a given E-cadherin gene even in those cells in which the mutant gene is expressed at a "normal" level, but the gene product is not fully functional.

Gene therapy would be carried out according to generally accepted methods, for example as described by Kren *et al.*, (1998), or as described by Friedman in *Therapy for Genetic Disease*, T. Friedman, ed., Oxford University Press (1991), pp 105-121. Cells from a patient would be first analyzed by the methods described above, to ascertain the production of E-cadherin polypeptide. A virus or plasmid vector, containing a copy of the E-cadherin gene linked to expression control elements and capable of replicating inside the target cells, is prepared. Suitable vectors are

known, such as disclosed in US Patent 5,252,479 and PCT published application WO 93/07282. The vector is then injected into the patient, either locally at the site of the target cells or systemically (in order to reach any target cells that may be at remote sites). If the transfected gene is not permanently incorporated into the genome of each of the targeted cells, the treatment may have to be repeated periodically.

Gene transfer systems known in the art may be useful in the practice of the gene therapy methods. These include viral and nonviral transfer methods. A number of viruses have been used as gene transfer vectors, including papovaviruses (eg. SV40, Madzak *et al.*, (1992)), adenovirus (Berkner (1992)), vaccinia virus (Moss (1992)), adeno-associated virus (Muzyczka (1992)), herpesviruses including HSV and EBV (Margolskee (1992); Johnson *et al.*, (1992); Fink *et al.*, (1992); Breakfield and Geller, (1987); Freese *et al.*, (1990)), and retroviruses of avian (Petropoulos *et al.*, (1992), murine (Miller (1992)); and human origin (Shimada *et al.*, (1991); Helseth *et al.*, (1990); Page *et al.*, (1990); Buchschacher and Panganiban (1992)).

Nonviral gene transfer methods known in the art include chemical techniques such as calcium phosphate coprecipitation (Pellicer *et al.*, (1980)); mechanical techniques, for example microinjection (Anderson *et al.*, (1980)); membrane fusion-mediated transfer via liposomes (Lim *et al.*, (1992)); and direct DNA uptake and receptor-mediated DNA transfer (Wolff *et al.*, (1990); Wu *et al.*, (1991)). Viral-mediated gene transfer can be combined with direct *in vivo* gene transfer using liposome delivery, allowing one to direct the viral vectors to the target cells. Alternatively, the retroviral vector producer cell line can be injected into the patient (Culver *et al.*, 1992). Injection of producer cells would then provide a continuous source of vector particles.

In an approach which combines biological and physical gene transfer methods, plasmid DNA of any size is combined with a polylysine-conjugated antibody specific to the adenovirus hexon protein, and the resulting complex is bound to an adenovirus vector. The trimolecular complex is then used to infect cells. The adenovirus vector permits efficient binding, internalization, and degradation of the endosome before the coupled DNA is damaged.

Liposome/DNA complexes have been shown to be capable of mediating direct *in vivo* gene transfer. While in standard liposome preparations the gene transfer process is nonspecific, localized *in vivo* uptake and expression have been reported in tumour deposits, for example, following direct *in situ* administration (Nabel, 1992).

Corrective efforts need not always involve gene therapy. Peptides which have wild-type E-cadherin activity can be supplied to cells which carry mutant or missing E-cadherin alleles as an alternative approach to gene therapy. Such peptides can be produced by expression of the cDNA sequence in bacteria, for example, using known expression vectors and known techniques (Sambrook *et al.*, (1989)). Alternatively, E-cadherin polypeptide can be extracted from E-cadherin-producing mammalian cells. In addition, the techniques of synthetic chemistry can be employed to synthesize E-cadherin protein (Merryfield, (1963)).

Active E-cadherin molecules can be introduced into cells by microinjection or by use of liposomes, for example. Alternatively, some active molecules may be taken up by cells, actively or by diffusion. Extracellular application of the E-cadherin gene product may be sufficient to prevent tumour growth. Supply of molecules with E-cadherin activity should lead to partial reversal of the risk of a later neoplastic state. Other molecules with E-cadherin activity (for example, peptides, drugs or organic compounds) may also be used to effect such a reversal. Modified polypeptides having substantially similar function can also be used for peptide therapy.

Still another implication of the applicant's finding is that cells which carry a mutant E-cadherin allele can be used as model systems to study and test for substances which have potential as prophylactic/therapeutic agents. The cells are typically cultured epithelial cells. These may be isolated from individuals with E-cadherin mutations, either somatic or germline. Alternatively, the cell line can be engineered to carry the mutation in the E-cadherin allele. After a test substance is applied to the cells, the neoplastically transformed phenotype of the cell is determined. Any trait of neoplastically transformed cells can be assessed, including anchorage-independent growth, tumourigenicity in nude mice, invasiveness of cells, and growth factor dependence. Assays for each of these traits are known in the art.

Those persons skilled in the art will appreciate that the above description is provided by way of example only and that it is limited only by the lawful scope of the appended claims.

REFERENCES

Grunwald, G. B. The structural and functional analysis of cadherin calcium-dependent cell adhesion molecules. *Curr. Opin. Cell Biol.* **5,** 797-805 (1993).

Shiozaki, H., Oka, H., Inoue, M., Tamura, S. & Monden, M. E-Cadherin mediated adhesion system in cancer cells. *Cancer* **77,** 1605-1613 (1995).

Bracke, M. E., Roy, F. M. & Mareel, M. M. The E Cadherin/Catenin complex in invasion and metastatsis. *Curr. Topics Micro. Imm.* **213**, 123-161 (1996).

Frixen, E. H. *et al.* E-Cadherin-mediated cell-cell adhesion prevents invasiveness of human carinoma cells. *J. Cell Biol.* **113,** 173-185 (1991).

Vleminckx, K., Vakaet, L., Mareel, M., Fiers, W. & Roy, F. V. Genetic manipulation of E-cadherin expression by epithelial tumour cells reveals an invasion suppressor role. *Cell* **66,** 107-119 (1991).

Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380,** 152-154 (1996).

Berx, G. *et al.* E-cadherin is a tumour/invasion suppressor gene mutated in human lobular breast cancers. *EMBO J.* **14,** 6107-6115 (1995).

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* **55,** 1119-1150 (1986).

Andrews, L. G. & Markert, M. L. Exon skipping in purine nucleoside phosphorylase mRNA processing leading to severe immunodeficiency. *J. Biol. Chem.* **267,** 7834-7838 (1992).

Kuivaniemi, H., Tromp, G., Bergfeld, W.F., Kay, M. & Helm, T.N. Ehlers-Danlos syndrome type IV: a single base substitution of the last nucleotide of exon 34 in COL3A1 leads to exon skipping. *J. Invest. Dermatol.* **105,** 352-356 (1995).

Oda, T. *et al.* E-cadherin gene mutations in human gastric carcinoma cell lines. *Proc. Natl. Acad. Sci. USA* **91,** 1858-1862 (1994).

Berx, G. *et al.* Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). *Genomics* **26,** 281-289 (1995).

Nagar, B., Overduin, M., Ikura, M. & Rini, J. M. Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature* **380,** 360-364 (1996).

Mahoney, P. A. *et al.* The fat tumour suppressor gene in Drosophila encodes a novel member of the cadherin gene superfamily. *Cell* **67,** 853-868 (1991).

Mareel, M., Bracke, M. & Van Roy, F. Cancer metastasis: negative regulation by an invasion-suppressor complex. *Cancer Detection Prevention* **19,** 451-464 (1995).

Banerjee, S.K., Makdisi, W.F., Weston, A.P., Mitchell, S.M. & Campbell, D.R. Microwave-based DNA extraction from paraffin-embedded tissue for PCR amplification. *Biotechniques* **18,** 768-773 (1995).

Chomczynski, P. & Sacchi, N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156-159 (1987).

Lathrop, G.M., Lalouel, J.M., Julier, C. & Ott, J. Multilocus linkage analysis in humans: Detection of linkage and estimation of recombination. *Am. J. hum. Genet.* **37,** 482-498 (1985).

H. Watanabe, J. R. Jass and L. H. Sobin. Histological Typing of Oesophageal and Gastric Tumours. 2nd ed. pp 20-24. Germany: Springer-Verlag Berlin Heidelberg, 1990.

Greer, C. E., Wheeler, C. M. and Manos, M. M. PCR amplification from paraffin-embedded tissues: sample preparation and the effects of fixation. *In:* C.W. Dieffenbach and G.S. Dveksler (eds) PCR primer. A laboratory manual. Cold Spring Harbor Laboratory Press 1995.

Grady, W. M., Rajput, A., Myeroff, L., Liu, D. F., Kwon, K.-H., Willis, J. and Markowitz, S. Mutation of the type II transforming growth factor-ß receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. Cancer Res., *58:* 3101-3104, 1998.

Berx, G., Cleton-Jansen, A.-M., Strumane, K., de Leeuw, W. J. F., Nollet, F., van Roy, F. and Cornelisse, C. E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. Oncogene, *13:* 1919-1925, 1996.

Berx, G., Staes, K., van Hengel, J., Molemans, F., Bussemakers, M. J. G., van Bokhoven, A. and van Roy F. Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1). Genomics, *26*: 281-289, 1995.

Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. and Sharp, P. A. Splicing of messenger RNA precursors. Ann. Rev. Biochem., *55*: 1119-1150, 1986.

Lauren, P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. Acta Pathol. Microbiol. Scand., *64:* 31-49, 1965.

Blaschuk, O. W., Sullivan, R., David, S. and Pouliot, Y. Identification of a cadherin cell adhesion recognition sequence. Dev. Biol., *139:* 227-229, 1990.

Anderson, *et al* (1980). *Proc. Natl. Acad. Sci.* USA 77:5399-5403.

Berkner (1992). *Curr. Top. Microbiol. Immunol.* 158:39-61.

Breakfield and Geller (1987). *Mol. Neurobiol.* 1:337-371.

Buchschacher and Panganiban (1992). *J. Virol.* 66:2731-2739.

Cariello (1988). *Human Genetics* 42:726.

Conner, B.J., *et al.* (1983). *Proc. Nat. Acad. Sci* USA. 80:278-282.

Culver, *et al.* (1992). *Science* 256:1550-1552.

Fink, *et al.* (1992). *Hum. Gene Ther.* 3:11-19.

Finkelstein, J., *et al.* (1990). *Genomics* 7:167-172.

Freese, *et al.* (1990). *Biochem. Pharmacol.* 40:2189-2199.

Goding (1986). *Monoclonal Antibodies: Principles and Practice,* 2d ed. (Academic Press, New York).

Harlow & Lane (1988). *Antibodies: A Laboratory Manual,* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

Helseth *et al,*. (1990). *J. Virol.* 64:2416-2420.

Huse, *et al,*. (1989). *Science* 246:1275-1281.

Innis *et al.,* (1990). *PCR Protocols: A Guide to Methods and Applications,* (Academic Press, San Diego).

Johnson, *et al.,* (1992). *J. Virol.* 66:2952-2965.

Kinszler, K.W., *et al.,* (1991). *Science* 251:1366-1370.

Kren *et al.,* (1998). *Nature Medicine*, 4, 285-290.

Lim, *et al.,* (1992). *Circulation* 83:2007-2011.

Madzak, *et al.*, (1992). *J. Gen. Virol.* 73:1533-1536.

Maniatis, T. *et al.,* (1982). *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

Margolskee (1992). *Curr. Top. Microbial. Immunol.* 158:67-90.

Merrifield (1963). *J. Am. Chem. Soc.* 85:2149-2156.

Modrich, P. (1991). *Ann. Rev. Genet.* 25:229-253.

Moss (1992). *Curr. Top. Microbiol. Immunol.* 158:25-38.

Muzyczka (1992). *Curr. Top. Microbiol. Immunol.* 158:97-123.

Newton, C.R., Graham, A., Heptinstall, L.E., Powell, S.J., Summers, C., Kalsheker, N., Smith, J.C., and Markham, A.F. (1989). *Nucl. Acids Res.* 17:2503-2516.

Novack, *et al.,* (1986). *Proc. Nat. Acad. Sci. USA* 83:586.

Orita, *et al.,* (1989). *Proc. Nat. Acad. Sci. USA* 86:276-2770.

Orita, *et al.,* (1989). *Genomics* 5:874-879.

Page, *et al.,* (1990). *J. Virol.* 64:5370-5276.

Petropoulos, *et al.*, (1992). *J. Virol.* 66:3391-3397.

Rano & Kidd (1989). *Nucl. Acids Res.* 17:8392.

Sambrook, J., *et al.*, (1989) *Molecular cloning: A laboratory manual*, 2nd Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

Scharf (186). *Science* 233:1076.

Sheffield, V.C., *et al.*, (1989). *Proc. Nat. Acad. Sci. USA* 86:232-236.

Shenk, *et al.*, (1975). *Proc. Nat. Acad. Sci. USA* 72:989.

Shimada, *et al.*, (1991). *J. Clin. Invest.* 88:1043-1047.

Wartell, R.M., *et al.*, (1990). *Nucl. Acids. Res.* 18:2699-2705.

Wolff, *et al.*, (1990). *Science* 247:1465-1468.

Wu, *et al.*, (1989a). *Genomics* 4:560-569.

Wu, *et al.*, (1991). *J. Biol. Chem.* 266:14338-14342.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: University of Otago
Te Wheta Whanau Trust Limited

(ii) TITLE OF INVENTION: GERMLINE MUTATIONS IN THE E-CADHERIN GENE AND METHOD FOR DETECTING PREDISPOSITIONS TO CANCER

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESS: Russell McVeagh West-Walker
(B) STREET: The Todd Building, Cnr Brandon Street and Lambton Quay
(C) CITY: Wellington
(D) COUNTRY: New Zealand

(v) COMPUTER READABLE FORM
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: Windows 95

(vi) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: NZ 328994
(B) FILING DATE: 17 October 1997

(vii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Bennett, Michael Roy
(B) REFERENCE/DOCKET NUMBER: 23677 MRB

(viii) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 64 4 499 9058
(B) TELEFAX: 64 4 499 9306

(2) INFORMATION FOR SEQ ID NO. 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4778 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 1:

gcttgcggaa gtcagttcag actccagccc gctccagccc ggcccgaccc gaccgcaccc 60
ggcgcctgcc ctcgctcggc gtccccggcc agccatgggc ccttggagcc gcagcctctc 120
ggcgctgctg ctgctgctgc aggtctcctc ttggctctgc caggagccgg agccctgcca 180
ccctggcttt gacgccgaga gctacacgtt cacggtgccc cggcgccacc tggagagagg 240
ccgcgtcctg ggcagagtga attttgaaga ttgcaccggt cgacaaagga cagcctattt 300
ttccctcgac acccgattca aagtgggcac agatggtgtg attacagtca aaaggcctct 360
acggtttcat aacccacaga tccatttctt ggtctacgcc tgggactcca cctacagaaa 420
gttttccacc aaagtcacgc tgaatacagt ggggcaccac caccgccccc cgccccatca 480
ggcctccgtt tctggaatcc aagcagaatt gctcacattt cccaactcct ctcctggcct 540
cagaagacag aagagagact gggttattcc tcccatcagc tgcccagaaa atgaaaaagg 600
cccatttcct aaaaacctgg ttcagatcaa atccaacaaa gacaaagaag gcaaggtttt 660
ctacagcatc actggccaag gagctgacac acccccctgtt ggtgtcttta ttattgaaag 720
agaaacagga tggctgaagg tgacagagcc tctggataga gaacgcattg ccacatacac 780
tctcttctct cacgctgtgt catccaacgg gaatgcagtt gaggatccaa tggagatttt 840
gatcacggta accgatcaga atgacaacaa gcccgaattc acccaggagg tctttaaggg 900
gtctgtcatg gaaggtgctc ttccaggaac ctctgtgatg gaggtcacag ccacagacgc 960
ggacgatgat gtgaacacct acaatgccgc catcgcttac accatcctca gccaagatcc 1020
tgagctccct gacaaaaata tgttcaccat taacaggaac acaggagtca tcagtgtggt 1080
caccactggg ctggaccgag agagtttccc tacgtatacc ctggtggttc aagctgctga 1140
ccttcaaggt gaggggttaa gcacaacagc aacagctgtg atcacagtca ctgacaccaa 1200
cgataatcct ccgatcttca atcccaccac gtacaagggt caggtgcctg agaacgaggc 1260
taacgtcgta atcaccacac tgaaagtgac tgatgctgat gcccccaata ccccagcgtg 1320
ggaggctgta tacaccatat tgaatgatga tggtggacaa tttgtcgtca ccacaaatcc 1380
agtgaacaac gatggcattt tgaaaacagc aaagggcttg gattttgagg ccaagcagca 1440
gtacattcta cacgtagcag tgacgaatgt ggtacctttt gaggtctctc tcaccacctc 1500
cacagccacc gtcaccgtgg atgtgctgga tgtgaatgaa gcccccatct ttgtgcctcc 1560
tgaaaagaga gtggaagtgt ccgaggactt tggcgtgggc caggaaatca catcctacac 1620
tgcccaggag ccagacacat ttatggaaca gaaaataaca tatcggattt ggagagacac 1680
tgccaactgg ctggagatta atccggacac tggtgccatt tccactcggg ctgagctgga 1740
cagggaggat tttgagcacg tgaagaacag cacgtacaca gccctaatca tagctacaga 1800
caatggttct ccagttgcta ctggaacagg gacacttctg ctgatcctgt ctgatgtgaa 1860
tgacaacgcc cccataccag aacctcgaac tatattcttc tgtgagagga atccaaagcc 1920
tcaggtcata aacatcattg atgcagacct tcctcccaat acatctccct tcacagcaga 1980

actaacacac ggggcgagtg ccaactggac cattcagtac aacgacccaa cccaagaatc 2040
tatcattttg aagccaaaga tggccttaga ggtgggtgac tacaaaatca atctcaagct 2100
catggataac cagaataaag accaagtgac caccttagag gtcagcgtgt gtgactgtga 2160
aggggccgcc ggcgtctgta ggaaggcaca gcctgtcgaa gcaggattgc aaattcctgc 2220
cattctgggg attcttggag gaattcttgc tttgctaatt ctgattctgc tgctcttgct 2280
gtttcttcgg aggagagcgg tggtcaaaga gcccttactg cccccagagg atgacacccg 2340
ggacaacgtt tattactatg atgaagaagg aggcggagaa gaggaccagg actttgactt 2400
gagccagctg cacaggggcc tggacgctcg gcctgaagtg actcgtaacg acgttgcacc 2460
aaccctcatg agtgtccccc ggtatcttcc ccgccctgcc aatcccgatg aaattggaaa 2520
ttttattgat gaaaatctga aagcggctga tactgacccc acagccccgc cttatgattc 2580
tctgctcgtg tttgactatg aaggaagcgg ttccgaagct gctagtctga gctccctgaa 2640
ctcctcagag tcagacaaag accaggacta tgactacttg aacgaatggg gcaatcgctt 2700
caagaagctg gctgacatgt acggaggcgg cgaggacgac taggggactc gagagaggcg 2760
ggccccagac ccatgtgctg ggaaatgcag aaatcacgtt gctggtggtt tttcagctcc 2820
cttcccttga gatgagtttc tggggaaaaa aaagagactg gttagtgatg cagttagtat 2880
agctttatac tctctccact ttatagctct aataagtttg tgttagaaaa gtttcgactt 2940
atttcttaaa gctttttttt ttttcccatc actctttaca tggtggtgat gtccaaaaga 3000
tacccaaatt ttaatattcc agaagaacaa ctttagcatc agaaggttca cccagcacct 3060
tgcagatttt cttaaggaat tttgtctcac ttttaaaaag aaggggagaa gtcagctact 3120
ctagttctgt tgttttgtgt atataatttt ttaaaaaaaa tttgtgtgct tctgctcatt 3180
actacactgg tgtgtccctc tgccttttt ttttttttta agacagggtc tcattctatc 3240
ggccaggctg gagtgcagtg gtgcaatcac agctcactgc agccttgtcc tcccaggctc 3300
aagctatcct tgcacctcag cctcccaagt agctgggacc acaggcatgc accactacgc 3360
atgactaatt ttttaaatat ttgagacggg gtctccctgt gttacccagg ctggtctcaa 3420
actcctgggc tcaagtgatc ctcccatctt ggcctcccag agtattggga ttacagacat 3480
gagccactgc acctgcccag ctccccaact ccctgccatt ttttaagaga cagtttcgct 3540
ccatcgccca ggcctgggat gcagtgatgt gatcatagct cactgtaacc tcaaactctg 3600
gggctcaagc agttctccca ccagcctcct ttttatttt ttgtacagat ggggtcttgc 3660
tatgttgccc aagctggtct taaactcctg gcctcaagca atccttctgc cttggccccc 3720
caaagtgctg ggattgtggg catgagctgc tgtgcccagc ctccatgttt taatatcaac 3780
tctcactcct gaattcagtt gctttgccca agataggagt tctctgatgc agaaattatt 3840
gggctctttt agggtaagaa gtttgtgtct ttgtctggcc acatcttgac taggtattgt 3900
ctactctgaa gacctttaat ggcttccctc tttcatctcc tgagtatgta acttgcaatg 3960
ggcagctatc cagtgacttg ttctgagtaa gtgtgttcat taatgtttat ttagctctga 4020

agcaagagtg atatactcca ggacttagaa tagtgcctaa agtgctgcag ccaaagacag 4080
agcggaacta tgaaaagtgg gcttggagat ggcaggagag cttgtcattg agcctggcaa 4140
tttagcaaac tgatgctgag gatgattgag gtgggtctac ctcatctctg aaaattctgg 4200
aaggaatgga ggagtctcaa catgtgtttc tgacacaaga tccgtggttt gtactcaaag 4260
cccagaatcc ccaagtgcct gcttttgatg atgtctacag aaaatgctgg ctgagctgaa 4320
cacatttgcc caattccagg tgtgcacaga aaaccgagaa tattcaaaat tccaaatttt 4380
ttcttaggag caagaagaaa atgtggccct aaaggggggtt agttgagggg taggggggtag 4440
tgaggatctt gatttggatc tctttttatt taaatgtgaa tttcaacttt tgacaatcaa 4500
agaaaagact tttgttgaaa tagctttact gtttctcaag tgttttggag aaaaaaatca 4560
accctgcaat cacttttttgg aattgtcttg atttttcggc agttcaagct atatcgaata 4620
tagttctgtg tagagaatgt cactgtagtt ttgagtgtat acatgtgtgg gtgctgataa 4680
ttgtgtattt tctttgggggg tggaaaagga aaacaattca agctgagaaa agtattctca 4740
aagatgcatt tttataaatt ttattaaaca attttgtt 4778

(2) INFORMATION FOR SEQ ID NO. 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 882 amino acids
(B) TYPE: amino acid
(C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 2:

Met Gly Pro Trp Ser Arg Ser Leu Ser Ala Leu Leu Leu Leu Leu Gln Val Ser
5 10 15
Ser Trp Leu Cys Gln Glu Pro Glu Pro Cys His Pro Gly Phe Asp Ala Glu Ser
20 25 30 35
Tyr Thr Phe Thr Val Pro Arg Arg His Leu Glu Arg Gly Arg Val Leu Gly Arg
40 45 50
Val Asn Phe Glu Asp Cys Thr Gly Arg Gln Arg Thr Ala Tyr Phe Ser Leu Asp
55 60 65 70
Thr Arg Phe Lys Val Gly Thr Asp Gly Val Ile Thr Val Lys Arg Pro Leu Arg
75 80 85 90
Phe His Asn Pro Gln Ile His Phe Leu Val Tyr Ala Trp Asp Ser Thr Tyr Arg
95 100 105
Lys Phe Ser Thr Lys Val Thr Leu Asn Thr Val Gly His His His Arg Pro Pro
110 115 120 125
Pro His Gln Ala Ser Val Ser Gly Ile Gln Ala Glu Leu Leu Thr Phe Pro Asn
130 135 140
Ser Ser Pro Gly Leu Arg Arg Gln Lys Arg Asp Trp Val Ile Pro Pro Ile Ser Cys
145 150 155 160

Pro Glu Asn Glu Lys Gly Pro Phe Pro Lys Asn Leu Val Gln Ile Lys Ser Asn
165 170 175 180
Lys Asp Lys Glu Gly Lys Val Phe Tyr Ser Ile Thr Gly Gln Gly Ala Asp Thr
185 190 195
Pro Pro Val Gly Val Phe Ile Ile Glu Arg Glu Thr Gly Trp Leu Lys Val Thr
200 205 210 215
Glu Pro Leu Asp Arg Glu Arg Ile Ala Thr Tyr Thr Leu Phe Ser His Ala Val
220 225 230 235
Ser Ser Asn Gly Asn Ala Val Glu Asp Pro Met Glu Ile Leu Ile Thr Val Thr
240 245 250
Asp Gln Asn Asp Asn Lys Pro Glu Phe Thr Gln Glu Val Phe Lys Gly Ser Val
255 260 265 270
Met Glu Gly Ala Leu Pro Gly Thr Ser Val Met Glu Val Thr Ala Thr Asp Ala
275 280 285
Asp Asp Asp Val Asn Thr Tyr Asn Ala Ala Ile Ala Tyr Thr Ile Leu Ser Gln
290 295 300 305
Asp Pro Glu Leu Pro Asp Lys Asn Met Phe Thr Ile Asn Arg Asn Thr Gly Val
310 315 320 325
Ile Ser Val Val Thr Thr Gly Leu Asp Arg Glu Ser Phe Pro Thr Tyr Thr Leu
330 335 340
Val Val Gln Ala Ala Asp Leu Gln Gly Glu Gly Leu Ser Thr Thr Ala Thr Ala
345 350 355 360
Val Ile Thr Val Thr Asp Thr Asn Asp Asn Pro Pro Ile Phe Asn Pro Thr Thr
365 370 375
Tyr Lys Gly Gln Val Pro Glu Asn Glu Ala Asn Val Val Ile Thr Thr Leu Lys
380 385 390 395
Val Thr Asp Ala Asp Ala Pro Asn Thr Pro Ala Trp Glu Ala Val Tyr Thr Ile
400 405 410 415
Leu Asn Asp Asp Gly Gly Gln Phe Val Val Thr Thr Asn Pro Val Asn Asn Asp
420 425 430
Gly Ile Leu Lys Thr Ala Lys Gly Leu Asp Phe Glu Ala Lys Gln Gln Tyr Ile
435 440 445 450
Leu His Val Ala Val Thr Asn Val Val Pro Phe Glu Val Ser Leu Thr Thr Ser
455 460 465
Thr Ala Thr Val Thr Val Asp Val Leu Asp Val Asn Glu Ala Pro Ile Phe Val
470 475 480 485
Pro Pro Glu Lys Arg Val Glu Val Ser Glu Asp Phe Gly Val Gly Gln Glu Ile
490 495 500 505
Thr Ser Tyr Thr Ala Gln Glu Pro Asp Thr Phe Met Glu Gln Lys Ile Thr Tyr
510 515 520
Arg Ile Trp Arg Asp Thr Ala Asn Trp Leu Glu Ile Asn Pro Asp Thr Gly Ala Ile
525 530 535 540
Ser Thr Arg Ala Glu Leu Asp Arg Glu Asp Phe Glu His Val Lys Asn Ser Thr
545 550 555 560
Tyr Thr Ala Leu Ile Ile Ala Thr Asp Asn Gly Ser Pro Val Ala Thr Gly Thr Gly
565 570 575
Thr Leu Leu Leu Ile Leu Ser Asp Val Asn Asp Asn Ala Pro Ile Pro Glu Pro
580 585 590 595
Arg Thr Ile Phe Phe Cys Glu Arg Asn Pro Lys Pro Gln Val Ile Asn Ile Ile Asp
600 605 610 615

Ala Asp Leu Pro Pro Asn Thr Ser Pro Phe Thr Ala Glu Leu Thr His Gly Ala
620 625 630
Ser Ala Asn Trp Thr Ile Gln Tyr Asn Asp Pro Thr Gln Glu Ser Ile Ile Leu Lys
635 640 645 650
Pro Lys Met Ala Leu Glu Val Gly Asp Tyr Lys Ile Asn Leu Lys Leu Met Asp
655 660 665 670
Asn Gln Asn Lys Asp Gln Val Thr Thr Leu Glu Val Ser Val Cys Asp Cys Glu
675 680 685
Gly Ala Ala Gly Val Cys Arg Lys Ala Gln Pro Val Glu Ala Gly Leu Gln Ile Pro
690 695 700 705
Ala Ile Leu Gly Ile Leu Gly Gly Ile Leu Ala Leu Leu Ile Leu Ile Leu Leu Leu
710 715 720 725
Leu Leu Phe Leu Arg Arg Arg Ala Val Val Lys Glu Pro Leu Leu Pro Pro Glu
730 735 740 745
Asp Asp Thr Arg Asp Asn Val Tyr Tyr Tyr Asp Glu Glu Gly Gly Gly Glu Glu
750 755 760
Asp Gln Asp Phe Asp Leu Ser Gln Leu His Arg Gly Leu Asp Ala Arg
765 770 775
Pro Glu Val Thr Arg Asn Asp Val Ala Pro Thr Leu Met Ser Val Pro Arg Tyr
780 785 790 795
Leu Pro Arg Pro Ala Asn Pro Asp Glu Ile Gly Asn Phe Ile Asp Glu Asn Leu
800 805 810 815
Lys Ala Ala Asp Thr Asp Pro Thr Ala Pro Pro Tyr Asp Ser Leu Leu Val Phe
820 825 830
Asp Tyr Glu Gly Ser Gly Ser Glu Ala Ala Ser Leu Ser Ser Leu Asn Ser Ser
835 840 845 850
Glu Ser Asp Lys Asp Gln Asp Tyr Asp Tyr Leu Asn Glu Trp Gly Asn Arg Phe
855 860 865
Lys Lys Leu Ala Asp Met Tyr Gly Gly Gly Glu Asp Asp
870 875 880

(2) INFORMATION FOR SEQ ID NO. 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 3:

taacaggaac acaggagtca tca

(2) INFORMATION FOR SEQ ID NO. 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 4:

gtggtgggat tgaagatcgg

(2) INFORMATION FOR SEQ ID NO. 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 5:

ttcccccacc ccaggtctc

(2) INFORMATION FOR SEQ ID NO. 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 6:

ccctcacctc tgcccaggac

(2) INFORMATION FOR SEQ ID NO. 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO. 7:

tgtagctctc ggcgtcaaag

CLAIMS

1. A method of testing to detect whether a human subject is predisposed to cancer which comprises the step of detecting the presence or absence of an alteration in the gene encoding E-cadherin, wherein the presence of an alteration is indicative of a predisposition to cancer.

2. A method for assessing a risk in a human subject for a predisposition for cancer which comprises the step of determining whether there is a germ-line alteration in the gene encoding E-cadherin, wherein the presence of an alteration is indicative of a risk for a predisposition for cancer.

3. A method according to claim 1 or claim 2 wherein presence or absence of an alteration is determined by analysis of DNA coding for E-cadherin.

4. A method according to claim 3 wherein the presence or absence of an alteration is determined by comparing the sequence of DNA from a sample from said subject with the DNA sequence coding for wild-type E-cadherin.

5. A method according to claim 1 or claim 2 wherein the presence or absence of an alteration is determined by analysis of mRNA transcribed from DNA encoding E-cadherin.

6. A method according to claim 5 wherein the presence or absence of an alteration is determined by comparing the sequence of mRNA from a sample from said subject with the mRNA sequence transcribed from DNA coding for wild-type E-cadherin.

7. A method according to claim 1 or claim 2 in which the presence or absence of an alteration is determined by analysis of the amino acid sequence of the expressed E-cadherin protein.

8. A method according to claim 7 wherein the presence or absence of an alteration is determined by comparing the amino acid sequence of the expressed E-cadherin protein from a sample from said subject with the amino acid sequence of wild-type E-cadherin protein.

9. A method according to claim 1 or claim 2 wherein the presence or absence of an alteration is determined by comparing the level of expression and/or activity of E-cadherin protein present in a sample from said subject with the level of expression/activity of wild-type E-cadherin protein.

10. A method according to claim 1 or claim 2 in which the presence of one or more of the following alterations in the gene encoding E-cadherin is indicative of a predisposition to cancer:

(i) G → T substitution at nucleotide 1008 (exon 7);

(ii) C insertion between nucleotides 2,382-2,386 (exon 15);

(iii) C → T substitution at nucleotide 2095 (exon 13);

(iv) C insertion at nucleotide 1588 (exon 11);

(v) G → T substitution at nucleotide 70 (exon 2); and

(vi) G → A substitution at nucleotide 1137 + 1 (donor splice site, intron 8).

11. A method according to any one of the preceding claims wherein the presence of an alteration is indicative of a predisposition, or a risk of predisposition, for gastric cancer.

12. A method according to claim 11 wherein the gastric cancer is hereditary diffuse gastric cancer (HDGC).

13. A method according to any one of claims 1 to 10 wherein the presence of an alteration is indicative of a predisposition, or a risk of predisposition, for colorectal cancer.

14. A method according to any one of claims 1 to 10 wherein the presence of an alteration is indicative of a predisposition, or a risk of predisposition, for breast cancer.

15. A method according to claim 10 in which the presence of one or more of alterations (i) to (vi) in the gene encoding E-cadherin is indicative of a predisposition to hereditary diffuse gastric cancer (HDGC).

16. A method of prophylactic and/or therapeutic treatment against cancer of an individual identified as having a risk of predisposition to cancer by a method according to any preceding claim which comprises the step of increasing, maintaining and/or restoring the active concentration of wild-type E-cadherin protein within said individual.

17. A method of prophylactic and/or therapeutic treatment against hereditary diffuse gastric cancer (HDGC) of an individual identified as having a risk of predisposition to cancer by a method according to claim 12 or claim 15 which comprises the step of increasing, maintaining and/or restoring the active concentration of wild-type E-cadherin protein within said individual.

18. A method of treatment according to claim 16 or claim 17 which comprises supplying said individual with wild-type E-cadherin gene function.

E-Cadherin mRNA coding Translated Sequence

Sequence Range: 1 to 2649

```
          10         20         30         40         50         60
           *          *          *          *          *          *
ATGGGCCCTT GGAGCCGCAG CCTCTCGGCG CTGCTGCTGC TGCTGCAGGT CTCCTCTTGG
MetGlyPro TrpSerArgSer LeuSerAla LeuLeuLeu LeuLeuGlnVal SerSerTrp>

          70         80         90        100        110        120
           *          *          *          *          *          *
CTCTGCCAGG AGCCGGAGCC CTGCCACCCT GGCTTTGACG CCGAGAGCTA CACGTTCACG
LeuCysGln GluProGluPro CysHisPro GlyPheAsp AlaGluSerTyr ThrPheThr>

         130        140        150        160        170        180
           *          *          *          *          *          *
GTGCCCCGGC GCCACCTGGA GAGAGGCCGC GTCCTGGGCA GAGTGAATTT TGAAGATTGC
ValProArg ArgHisLeuGlu ArgGlyArg ValLeuGly ArgValAsnPhe GluAspCys>

         190        200        210        220        230        240
           *          *          *          *          *          *
ACCGGTCGAC AAAGGACAGC CTATTTTTCC CTCGACACCC GATTCAAAGT GGGCACAGAT
ThrGlyArg GlnArgThrAla TyrPheSer LeuAspThr ArgPheLysVal GlyThrAsp>

         250        260        270        280        290        300
           *          *          *          *          *          *
GGTGTGATTA CAGTCAAAAG GCCTCTACGG TTTCATAACC CACAGATCCA TTTCTTGGTC
GlyValIle ThrValLysArg ProLeuArg PheHisAsn ProGlnIleHis PheLeuVal>

         310        320        330        340        350        360
           *          *          *          *          *          *
TACGCCTGGG ACTCCACCTA CAGAAAGTTT TCCACCAAAG TCACGCTGAA TACAGTGGGG
TyrAlaTrp AspSerThrTyr ArgLysPhe SerThrLys ValThrLeuAsn ThrValGly>

         370        380        390        400        410        420
           *          *          *          *          *          *
CACCACCACC GCCCCCCGCC CCATCAGGCC TCCGTTTCTG GAATCCAAGC AGAATTGCTC
HisHisHis ArgProProPro HisGlnAla SerValSer GlyIleGlnAla GluLeuLeu>

         430        440        450        460        470        480
           *          *          *          *          *          *
ACATTTCCCA ACTCCTCTCC TGGCCTCAGA AGACAGAAGA GAGACTGGGT TATTCCTCCC
ThrPhePro AsnSerSerPro GlyLeuArg ArgGlnLys ArgAspTrpVal IleProPro>

         490        500        510        520        530        540
           *          *          *          *          *          *
ATCAGCTGCC CAGAAAATGA AAAAGGCCCA TTTCCTAAAA ACCTGGTTCA GATCAAATCC
IleSerCys ProGluAsnGlu LysGlyPro PheProLys AsnLeuValGln IleLysSer>

         550        560        570        580        590        600
           *          *          *          *          *          *
AACAAAGACA AAGAAGGCAA GGTTTTCTAC AGCATCACTG GCCAAGGAGC TGACACACCC
AsnLysAsp LysGluGlyLys ValPheTyr SerIleThr GlyGlnGlyAla AspThrPro>

         610        620        630        640        650        660
           *          *          *          *          *          *
CCTGTTGGTG TCTTTATTAT TGAAAGAGAA ACAGGATGGC TGAAGGTGAC AGAGCCTCTG
ProValGly ValPheIleIle GluArgGlu ThrGlyTrp LeuLysValThr GluProLeu>

         670        680        690        700        710        720
           *          *          *          *          *          *
GATAGAGAAC GCATTGCCAC ATACACTCTC TTCTCTCACG CTGTGTCATC CAACGGGAAT
AspArgGlu ArgIleAlaThr TyrThrLeu PheSerHis AlaValSerSer AsnGlyAsn>

         730        740        750        760        770        780
           *          *          *          *          *          *
GCAGTTGAGG ATCCAATGGA GATTTTGATC ACGGTAACCG ATCAGAATGA CAACAAGCCC
AlaValGlu AspProMetGlu IleLeuIle ThrValThr AspGlnAsnAsp AsnLysPro>
```

FIG 1

E-Cadherin mRNA coding Translated Sequence

```
       790        800        810        820        830        840
        *          *          *          *          *          *
GAATTCACCC AGGAGGTCTT TAAGGGGTCT GTCATGGAAG GTGCTCTTCC AGGAACCTCT
GluPheThr GlnGluValPhe LysGlySer ValMetGlu GlyAlaLeuPro GlyThrSer>

       850        860        870        880        890        900
        *          *          *          *          *          *
GTGATGGAGG TCACAGCCAC AGACGCGGAC GATGATGTGA ACACCTACAA TGCCGCCATC
ValMetGlu ValThrAlaThr AspAlaAsp AspAspVal AsnThrTyrAsn AlaAlaIle>

       910        920        930        940        950        960
        *          *          *          *          *          *
GCTTACACCA TCCTCAGCCA AGATCCTGAG CTCCCTGACA AAAATATGTT CACCATTAAC
AlaTyrThr IleLeuSerGln AspProGlu LeuProAsp LysAsnMetPhe ThrIleAsn>

       970        980        990       1000       1010       1020
        *          *          *          *          *          *
AGGAACACAG GAGTCATCAG TGTGGTCACC ACTGGGCTGG ACCGAGAGAG TTTCCCTACG
ArgAsnThr GlyValIleSer ValValThr ThrGlyLeu AspArgGluSer PheProThr>

      1030       1040       1050       1060       1070       1080
        *          *          *          *          *          *
TATACCCTGG TGGTTCAAGC TGCTGACCTT CAAGGTGAGG GGTTAAGCAC AACAGCAACA
TyrThrLeu ValValGlnAla AlaAspLeu GlnGlyGlu GlyLeuSerThr ThrAlaThr>

      1090       1100       1110       1120       1130       1140
        *          *          *          *          *          *
GCTGTGATCA CAGTCACTGA CACCAACGAT AATCCTCCGA TCTTCAATCC CACCACGTAC
AlaValIle ThrValThrAsp ThrAsnAsp AsnProPro IlePheAsnPro ThrThrTyr>

      1150       1160       1170       1180       1190       1200
        *          *          *          *          *          *
AAGGGTCAGG TGCCTGAGAA CGAGGCTAAC GTCGTAATCA CCACACTGAA AGTGACTGAT
LysGlyGln ValProGluAsn GluAlaAsn ValValIle ThrThrLeuLys ValThrAsp>

      1210       1220       1230       1240       1250       1260
        *          *          *          *          *          *
GCTGATGCCC CCAATACCCC AGCGTGGGAG GCTGTATACA CCATATTGAA TGATGATGGT
AlaAspAla ProAsnThrPro AlaTrpGlu AlaValTyr ThrIleLeuAsn AspAspGly>

      1270       1280       1290       1300       1310       1320
        *          *          *          *          *          *
GGACAATTTG TCGTCACCAC AAATCCAGTG AACAACGATG GCATTTTGAA AACAGCAAAG
GlyGlnPhe ValValThrThr AsnProVal AsnAsnAsp GlyIleLeuLys ThrAlaLys>

      1330       1340       1350       1360       1370       1380
        *          *          *          *          *          *
GGCTTGGATT TTGAGGCCAA GCAGCAGTAC ATTCTACACG TAGCAGTGAC GAATGTGGTA
GlyLeuAsp PheGluAlaLys GlnGlnTyr IleLeuHis ValAlaValThr AsnValVal>

      1390       1400       1410       1420       1430       1440
        *          *          *          *          *          *
CCTTTTGAGG TCTCTCTCAC CACCTCCACA GCCACCGTCA CCGTGGATGT GCTGGATGTG
ProPheGlu ValSerLeuThr ThrSerThr AlaThrVal ThrValAspVal LeuAspVal>

      1450       1460       1470       1480       1490       1500
        *          *          *          *          *          *
AATGAAGCCC CCATCTTTGT GCCTCCTGAA AAGAGAGTGG AAGTGTCCGA GGACTTTGGC
AsnGluAla ProIlePheVal ProProGlu LysArgVal GluValSerGlu AspPheGly>

      1510       1520       1530       1540       1550       1560
        *          *          *          *          *          *
GTGGGCCAGG AAATCACATC CTACACTGCC CAGGAGCCAG ACACATTTAT GGAACAGAAA
ValGlyGln GluIleThrSer TyrThrAla GlnGluPro AspThrPheMet GluGlnLys>
```

FIG 1 (cont'd)

E-Cadherin mRNA coding Translated Sequence

```
      1570        1580        1590        1600        1610        1620
        *           *           *           *           *           *
ATAACATATC GGATTTGGAG AGACACTGCC AACTGGCTGG AGATTAATCC GGACACTGGT
IleThrTyr ArgIleTrpArg AspThrAla AsnTrpLeu GluIleAsnPro AspThrGly>

      1630        1640        1650        1660        1670        1680
        *           *           *           *           *           *
GCCATTTCCA CTCGGGCTGA GCTGGACAGG GAGGATTTTG AGCACGTGAA CAACAGCACG
AlaIleSer ThrArgAlaGlu LeuAspArg GluAspPhe GluHisValLys AsnSerThr>

      1690        1700        1710        1720        1730        1740
        *           *           *           *           *           *
TACACAGCCC TAATCATAGC TACAGACAAT GGTTCTCCAG TTGCTACTGG AACAGGGACA
TyrThrAla LeuIleIleAla ThrAspAsn GlySerPro ValAlaThrGly ThrGlyThr>

      1750        1760        1770        1780        1790        1800
        *           *           *           *           *           *
CTTCTGCTGA TCCTGTCTGA TGTGAATGAC AACGCCCCCA TACCAGAACC TCGAACTATA
LeuLeuLeu IleLeuSerAsp ValAsnAsp AsnAlaPro IleProGluPro ArgThrIle>

      1810        1820        1830        1840        1850        1860
        *           *           *           *           *           *
TTCTTCTGTG AGAGGAATCC AAAGCCTCAG GTCATAAACA TCATTGATGC AGACCTTCCT
PhePheCys GluArgAsnPro LysProGln ValIleAsn IleIleAspAla AspLeuPro>

      1870        1880        1890        1900        1910        1920
        *           *           *           *           *           *
CCCAATACAT CTCCCTTCAC AGCAGAACTA ACACACGGGG CGAGTGCCAA CTGGACCATT
ProAsnThr SerProPheThr AlaGluLeu ThrHisGly AlaSerAlaAsn TrpThrIle>

      1930        1940        1950        1960        1970        1980
        *           *           *           *           *           *
CAGTACAACG ACCCAACCCA AGAATCTATC ATTTTGAAGC CAAAGATGGC CTTAGAGGTG
GlnTyrAsn AspProThrGln GluSerIle IleLeuLys ProLysMetAla LeuGluVal>

      1990        2000        2010        2020        2030        2040
        *           *           *           *           *           *
GGTGACTACA AAATCAATCT CAAGCTCATG GATAACCAGA ATAAAGACCA AGTGACCACC
GlyAspTyr LysIleAsnLeu LysLeuMet AspAsnGln AsnLysAspGln ValThrThr>

      2050        2060        2070        2080        2090        2100
        *           *           *           *           *           *
TTAGAGGTCA GCGTGTGTGA CTGTGAAGGG GCCGCCGGCG TCTGTAGGAA GGCACAGCCT
LeuGluVal SerValCysAsp CysGluGly AlaAlaGly ValCysArgLys AlaGlnPro>

      2110        2120        2130        2140        2150        2160
        *           *           *           *           *           *
GTCGAAGCAG GATTGCAAAT TCCTGCCATT CTGGGGATTC TTGGAGGAAT TCTTGCTTTG
ValGluAla GlyLeuGlnIle ProAlaIle LeuGlyIle LeuGlyGlyIle LeuAlaLeu>

      2170        2180        2190        2200        2210        2220
        *           *           *           *           *           *
CTAATTCTGA TTCTGCTGCT CTTGCTGTTT CTTCGGAGGA GAGCGGTGGT CAAAGAGCCC
LeuIleLeu IleLeuLeuLeu LeuLeuPhe LeuArgArg ArgAlaValVal LysGluPro>

      2230        2240        2250        2260        2270        2280
        *           *           *           *           *           *
TTACTGCCCC CAGAGGATGA CACCCGGGAC AACGTTTATT ACTATGATGA ACAAGGAGGC
LeuLeuPro ProGluAspAsp ThrArgAsp AsnValTyr TyrTyrAspGlu GluGlyGly>

      2290        2300        2310        2320        2330        2340
        *           *           *           *           *           *
GGAGAAGAGG ACCAGGACTT TGACTTGAGC CAGCTGCACA GGGGCCTGGA CGCTCGGCCT
GlyGluGlu AspGlnAspPhe AspLeuSer GlnLeuHis ArgGlyLeuAsp AlaArgPro>

      2350        2360        2370        2380        2390        2400
```

FIG 1 (cont'd)

E-Cadherin mRNA coding Translated Sequence

```
          *           *           *           *           *           *
GAAGTGACTC GTAACGACGT TGCACCAACC CTCATGAGTG TCCCCCGGTA TCTTCCCCGC
GluValThr ArgAsnAspVal AlaProThr LeuMetSer ValProArgTyr LeuProArg>

      2410        2420        2430        2440        2450        2460
         *           *           *           *           *           *
CCTGCCAATC CCGATGAAAT TGGAAATTTT ATTGATGAAA ATCTGAAAGC GGCTGATACT
ProAlaAsn ProAspGluIle GlyAsnPhe IleAspGlu AsnLeuLysAla AlaAspThr>

      2470        2480        2490        2500        2510        2520
         *           *           *           *           *           *
GACCCCACAG CCCCGCCTTA TGATTCTCTG CTCGTGTTTG ACTATGAAGG AAGCGGTTCC
AspProThr AlaProProTyr AspSerLeu LeuValPhe AspTyrGluGly SerGlySer>

      2530        2540        2550        2560        2570        2580
         *           *           *           *           *           *
GAAGCTGCTA GTCTGAGCTC CCTGAACTCC TCAGAGTCAG ACAAAGACCA GGACTATGAC
GluAlaAla SerLeuSerSer LeuAsnSer SerGluSer AspLysAspGln AspTyrAsp>

      2590        2600        2610        2620        2630        2640
         *           *           *           *           *           *
TACTTGAACG AATGGGGCAA TCGCTTCAAG AAGCTGGCTG ACATGTACGG AGGCGGCGAG
TyrLeuAsn GluTrpGlyAsn ArgPheLys LysLeuAla AspMetTyrGly GlyGlyGlu>

GACGACTAG
AspAsp***>
```

FIG 1 (cont'd)

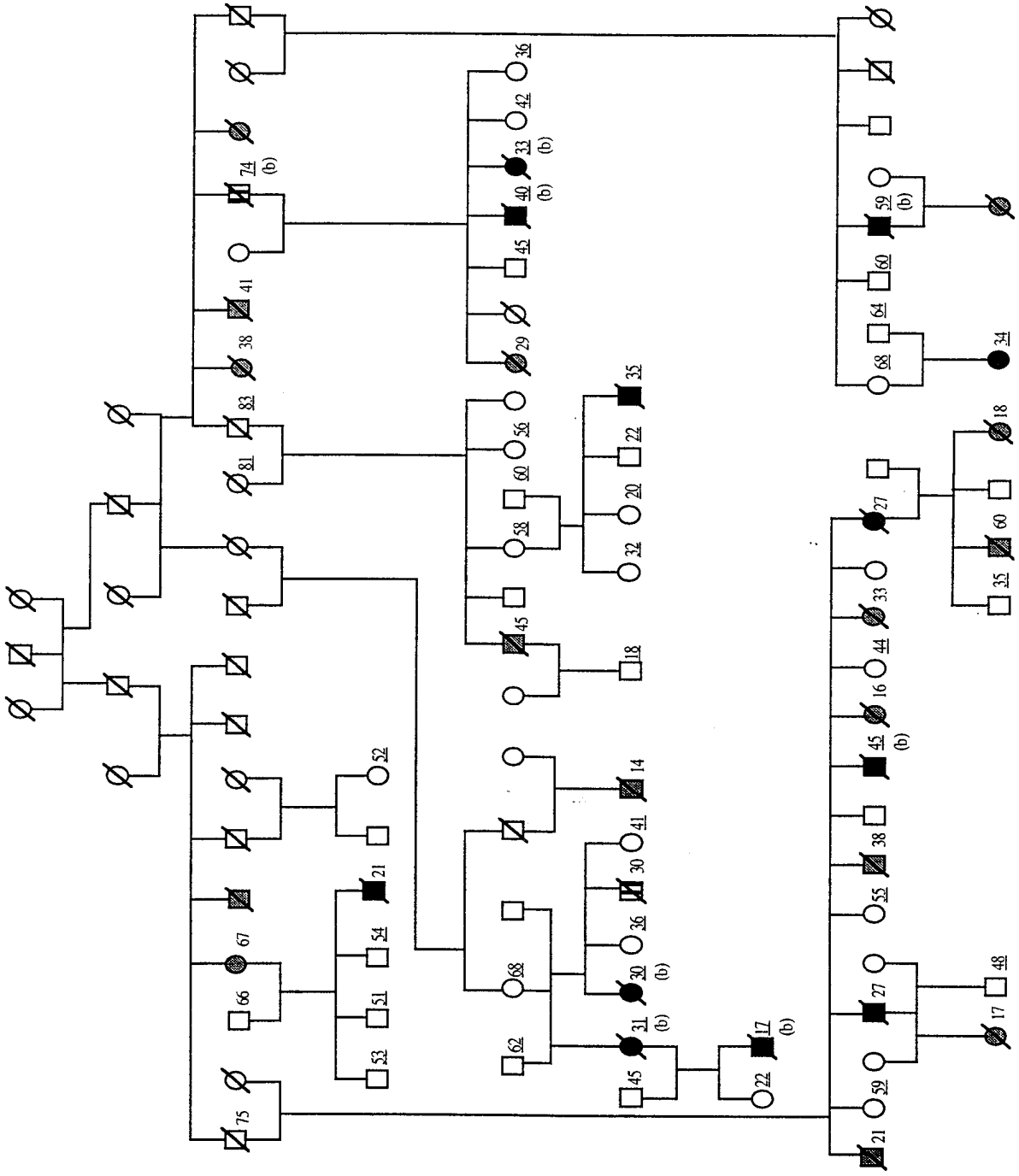


FIG 2

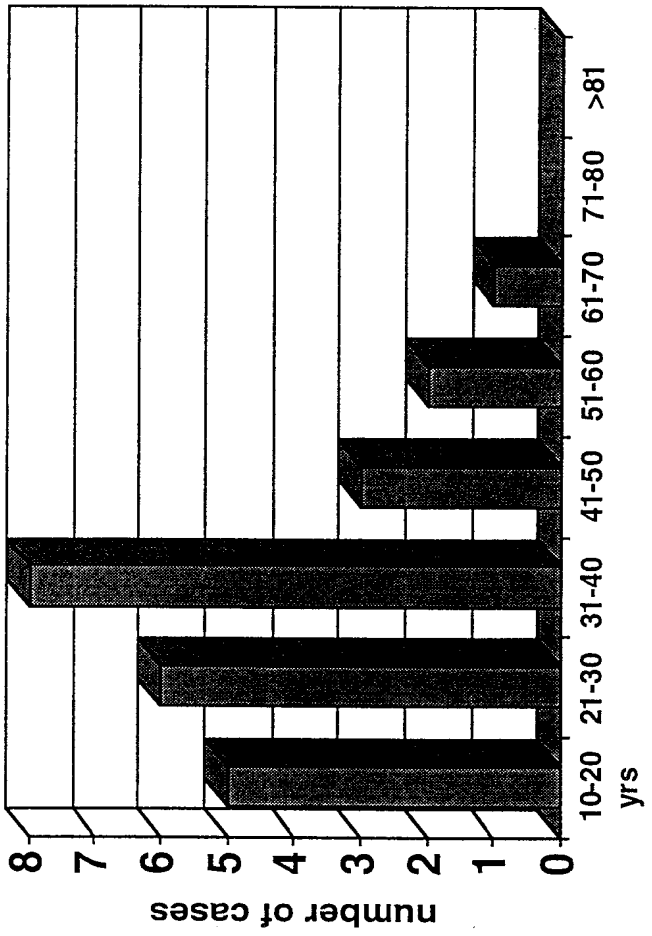

number of cases
0
1
2
3
4
5
6
7
8
10-20
21-30
31-40
41-50
51-60
61-70
71-80
>81
yrs

FIG 3

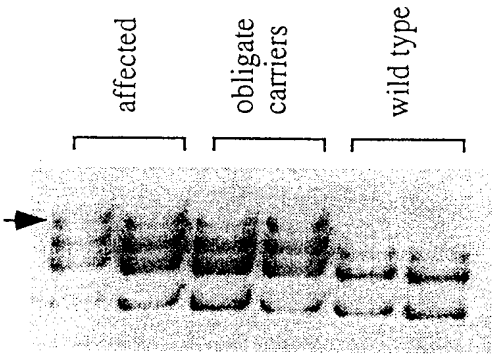

affected
obligate carriers
wild type

FIG 4a

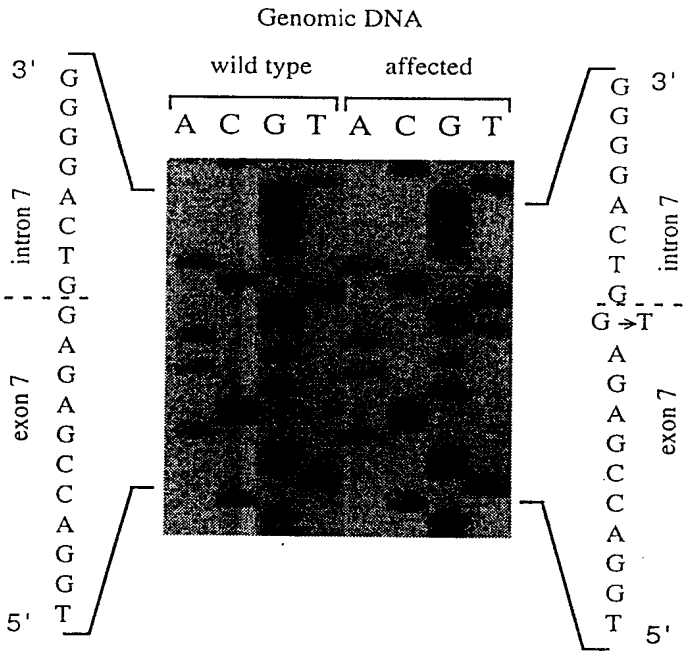

Genomic DNA
wild type
affected
A C G T A C G T
3'
5'
intron 7
exon 7
G G G G A C T G
G A G A G C C A G G T
G→T

FIG 4b

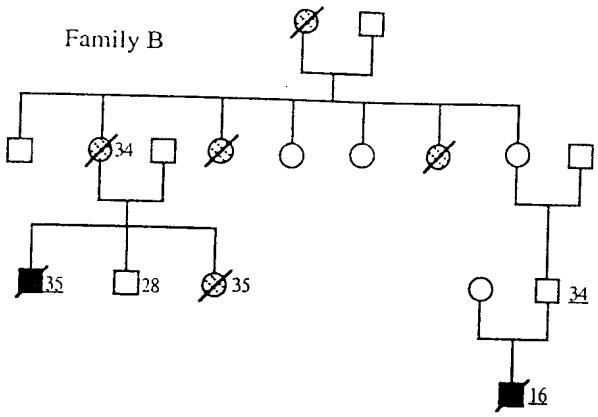


FIG 5

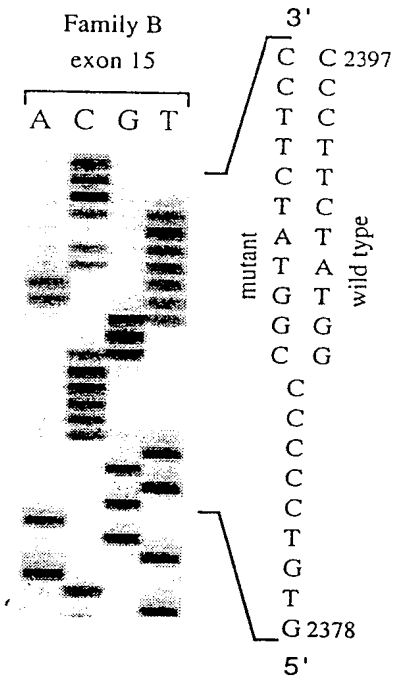

Family B
exon 15
A C G T
3'
C C 2397
C C
T C
T T
C T
T C
A T
T A
G T
G G
C G
C
C
C
C
C
T
G
T
G 2378
5'
mutant
wild type

FIG 6A

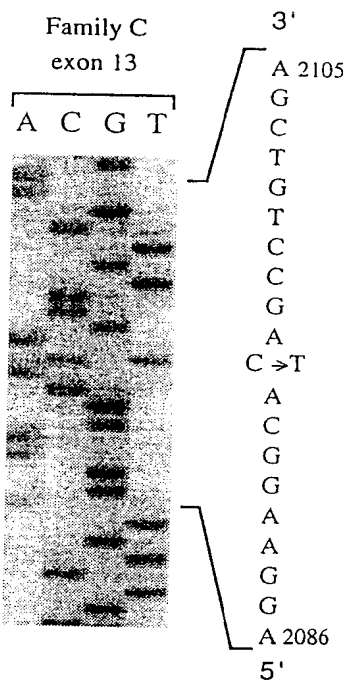
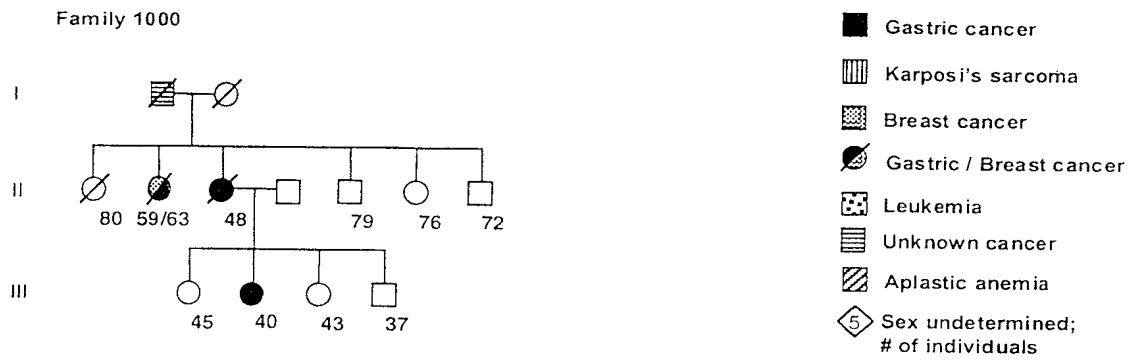
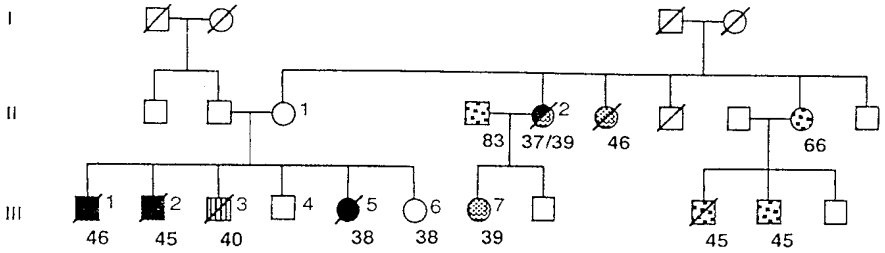

Family C
exon 13
A C G T
3'
A 2105
G
C
T
G
T
C
C
G
A
C →T
A
C
G
G
A
A
G
G
A 2086
5'

FIG 6B

a.


Family 1000
I
II
III
80
59/63
48
79
76
72
45
40
43
37
Gastric cancer
Karposi's sarcoma
Breast cancer
Gastric / Breast cancer
Leukemia
Unknown cancer
Aplastic anemia
5 Sex undetermined; # of individuals
12 Patient #
38 Age at diagnosis, or present age of unaffected individuals

b.


Family 4201
I
II
III
1
2
83
37/39
46
66
1
46
2
45
3
40
4
5
38
6
38
7
39
45
45

c.

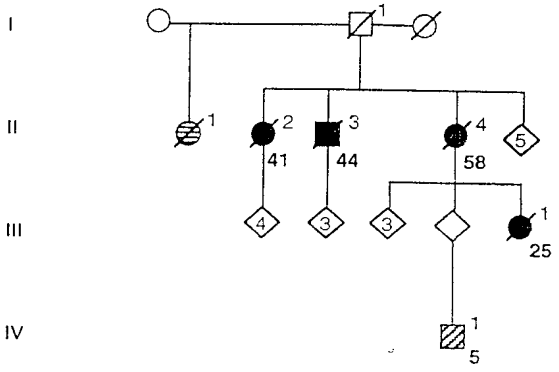

Family CHG 72
I
II
III
IV
1
1
2
41
3
44
4
58
5
4
3
3
1
25
1
5

FIG 7

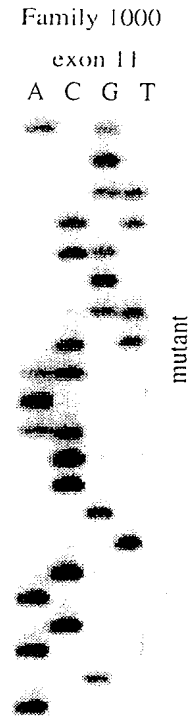


FIG 8A

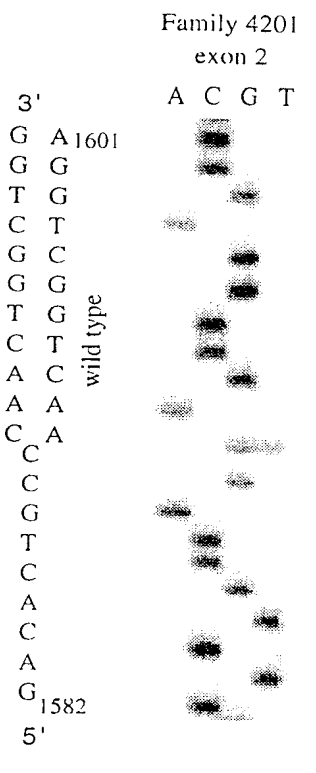


FIG 8B

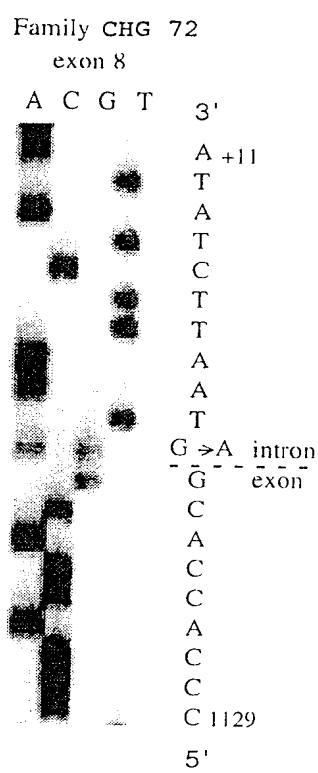


FIG 8C